(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 818 622**

(51) Int. CI.:

H04N 21/854 (2011.01)
H04N 19/30 (2014.01)
H04N 21/2343 (2011.01)
H04N 21/845 (2011.01)
H04N 19/70 (2014.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **24.05.2017 PCT/US2017/034290**

(87) Fecha y número de publicación internacional: **30.11.2017 WO17205518**

(96) Fecha de presentación y número de la solicitud europea: **24.05.2017 E 17729604 (3)**

(97) Fecha y número de publicación de la concesión europea: **17.06.2020 EP 3466095**

(54) Título: **Entradas de muestra y acceso aleatorio**

(30) Prioridad:

24.05.2016 US 201662340886 P
23.05.2017 US 201715602988

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
13.04.2021

(73) Titular/es:

QUALCOMM INCORPORATED (100.0%)
5775 Morehouse Drive
San Diego, CA 92121-1714, US

(72) Inventor/es:

WANG, YE-KUI

(74) Agente/Representante:

FORTEA LAGUNA, Juan José

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Entradas de muestra y acceso aleatorio

[0001] Esta solicitud reivindica el beneficio de la solicitud provisional estadounidense número 62/340.886**,** presentada el 24 de mayo de 2016.

CAMPO TÉCNICO

[0002] Esta divulgación se refiere a formatos de archivo para el almacenamiento y transporte de flujos de bits de medios, como los flujos de bits de vídeo.

ANTECEDENTES

[0003] Las capacidades de vídeo digital se pueden incorporar a una amplia gama de dispositivos, incluyendo televisores digitales, sistemas de radiodifusión directa digital, sistemas de radiodifusión inalámbrica, asistentes personales digitales (PDA), ordenadores portátiles o de escritorio, cámaras digitales, dispositivos de grabación digitales, reproductores de medios digitales, dispositivos de videojuegos, consolas de videojuegos, teléfonos de radio celulares o por satélite, dispositivos de videoconferencia y similares. Los dispositivos de vídeo digitales implementan técnicas de compresión de vídeo, tales como las descritas en las normas definidas por MPEG-2, MPEG-4, ITU-T H.263 o ITU-T H.264/MPEG-4, parte 10, Codificación de Vídeo Avanzada (AVC) y ampliaciones de dichas normas, para transmitir y recibir información de vídeo digital de manera más eficaz.

[0004] Las técnicas de compresión de vídeo realizan predicción espacial y/o predicción temporal para reducir o eliminar la redundancia inherente a las secuencias de vídeo. Para la codificación de vídeo basada en bloques, una trama o un fragmento de vídeo se pueden dividir en macrobloques. Cada macrobloque se puede dividir aún más. Los macrobloques de una trama o un fragmento intracodificados (I) se codifican usando predicción espacial con respecto a macrobloques vecinos. Los macrobloques de una trama o un fragmento intercodificados (P o B) pueden usar predicción espacial con respecto a macrobloques vecinos de la misma trama o fragmento, o predicción temporal con respecto a otras tramas de referencia.

[0005] Después de que se hayan codificado los datos de vídeo, los datos de vídeo se pueden agrupar en paquetes para su transmisión o almacenamiento. Los datos de vídeo se pueden ensamblar en un archivo de vídeo que se ajusta a cualquiera de una variedad de normas, tales como el formato de archivo de medios de base de la Organización Internacional de Normalización (ISO) y ampliaciones del mismo, tal como el formato de archivo AVC.

BREVE EXPLICACIÓN

[0006] En general, esta divulgación describe técnicas para permitir operaciones de acceso aleatorio convenientes para archivos que almacenan un flujo de bits de una o varias capas en una o más pistas. Esta divulgación proporciona procedimientos, dispositivos y productos de programa informático en diseños de entrada de muestras que permiten operaciones de acceso aleatorio convenientes, para archivos que almacenan un flujo de bits de una o varias capas en una o más pistas.

[0007] Los detalles de uno o más ejemplos se exponen en los dibujos adjuntos y en la siguiente descripción. Otras características, objetivos y ventajas resultarán evidentes a partir de la descripción y de los dibujos, y a partir de las reivindicaciones.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

[0008]

La FIG. 1 es un diagrama de bloques que ilustra un sistema de ejemplo que implementa técnicas para transmitir en continuo datos de medios a través de una red.

La FIG. 2 es un diagrama de bloques que ilustra un conjunto de componentes de ejemplo de una unidad de recuperación.

La FIG. 3 es un diagrama conceptual que ilustra elementos de contenido multimedia de ejemplo.

La FIG. 4 es un diagrama de bloques que ilustra elementos de un archivo de vídeo de ejemplo.

La FIG. 5 es un diagrama de flujo que ilustra una técnica de ejemplo para procesar datos.

La FIG. 6 es un diagrama de flujo que ilustra una técnica de ejemplo para procesar datos.

La FIG. 7 es un diagrama de flujo que ilustra una técnica de ejemplo para generar un archivo que incluye datos de vídeo.

DESCRIPCIÓN DETALLADA

[0009] En general, esta divulgación describe técnicas para permitir operaciones de acceso aleatorio convenientes para archivos que almacenan un flujo de bits de una o varias capas en una o más pistas. El acceso aleatorio conveniente puede definirse como el acceso aleatorio sin necesidad de buscar y/o recuperar conjuntos de parámetros de muestras anteriores. En otras palabras, se habilita el acceso aleatorio conveniente para una muestra de un flujo de bits de vídeo cuando un dispositivo cliente puede solicitar la muestra y muestras posteriores sin buscar y/o recuperar conjuntos de parámetros de muestras anteriores. Por lo tanto, todos los conjuntos de parámetros necesarios pueden incluirse en la propia muestra o en una entrada de muestra que corresponda a la muestra.

[0010] Esta divulgación describe procedimientos en diseños de entrada de muestra que permiten operaciones de acceso aleatorio convenientes, para archivos que almacenan un flujo de bits de una o varias capas en una o más pistas.

[0011] Las normas de codificación de vídeo incluyen ITU-T H.261, ISO/IEC MPEG-1 Visual, ITU-T H.262 o ISO/IEC MPEG-2 Visual, ITU-T H.263, ISO/IEC MPEG-4 Visual, ITU-T H.264 o ISO/IEC MPEG-4 AVC, incluyendo sus ampliaciones de codificación de vídeo escalable (SVC) y codificación de vídeo multivista (MVC), y codificación de vídeo de alta eficacia (HEVC), también conocida como ITU-T H.265 e ISO/IEC 23008-2, incluyendo su ampliación de codificación escalable (es decir, codificación de vídeo escalable de alta eficacia, SHVC), su ampliación multivista (es decir, codificación de vídeo multivista de alta eficacia, MV-HEVC) y su ampliación 3D (es decir, codificación de vídeo 3D de alta eficacia, 3D-HEVC).

[0012] Las normas de formato de archivo incluyen el formato de archivo de medios de base ISO (ISOBMFF, ISO/IEC 14496-12) y otros derivados del ISOBMFF, incluido el formato de archivo MPEG-4 (ISO/IEC 14496-15), el formato de archivo 3GPP (3GPP TS 26.244) e ISO/IEC 14496-15 que contiene los formatos de archivo para AVC y sus ampliaciones, así como los formatos de archivo para HEVC y sus ampliaciones. Los borradores de las nuevas ediciones recientes para ISO/IEC 14496-12 y 14496-15 están disponibles en http://phenix.intevry.fr/mpeg/doc_end_user/documents/111_Geneva/wg11/w15177-v6-w15177.zip y http://wg11.sc29.org/doc_end_user/documents/114_San%20Diego/wg11/w15928-v2-w15928.zip, respectivamente.

[0013] Las técnicas de esta divulgación se pueden aplicar a los archivos de vídeo conformes a los datos de vídeo encapsulados de acuerdo con cualquiera de entre el formato de archivo de medios de base ISO, el formato de archivo de codificación de vídeo escalable (SVC), el formato de archivo de codificación de vídeo avanzada (AVC), el formato de archivo del Proyecto de colaboración de tercera generación (3GPP) y/o el formato de archivo de codificación de vídeo de múltiples vistas (MVC) u otros formatos similares de archivo de vídeo.

[0014] En la transmisión continua HTTP, las operaciones de uso frecuente incluyen HEAD, GET y GET parcial. La operación HEAD recupera una cabecera de un archivo asociado a un localizador de recursos uniforme (URL) o un nombre de recursos uniforme (URN) dado, sin recuperar una carga útil asociada al URL o al URN. La operación GET recupera un archivo entero asociado a un URL o URN dado. La operación GET parcial recibe un intervalo de bytes como parámetro de entrada y recupera un número continuo de bytes de un archivo, donde el número de bytes corresponde al intervalo de bytes recibido. Por tanto, se pueden proporcionar fragmentos de película para la transmisión continua del HTTP, porque una operación GET parcial puede obtener uno o más fragmentos de película individuales. En un fragmento de película, pueden existir varios fragmentos de pista de diferentes pistas. En la transmisión continua del HTTP, una presentación de medios puede ser una recopilación de datos estructurados que es accesible para el cliente. El cliente puede solicitar y descargar la información de datos de medios para presentar un servicio de transmisión continua a un usuario.

[0015] En el ejemplo de transmisión en continuo de datos 3GPP usando transmisión en continuo HTTP, pueden existir múltiples representaciones para datos de vídeo y/o audio de contenido multimedia. Como se describe a continuación, diferentes representaciones pueden corresponder a diferentes características de codificación (por ejemplo, diferentes perfiles o niveles de una norma de codificación de vídeo), diferentes normas de codificación o ampliaciones de normas de codificación (tales como ampliaciones multivista y/o escalables), o diferentes tasas de bits. El manifiesto de dichas representaciones se puede definir en una estructura de datos de descripción de presentación de medios (MPD). Una presentación de medios puede corresponder a un grupo de datos estructurado que es accesible para un dispositivo cliente de transmisión en continuo HTTP. El dispositivo cliente de transmisión en continuo HTTP puede solicitar y descargar información de datos de medios para presentar un servicio de transmisión en continuo a un usuario del dispositivo cliente. Una presentación de medios se puede describir en la estructura de datos MPD, que puede incluir actualizaciones de la MPD.

[0016] Una presentación de medios puede contener una secuencia de uno o más períodos. Cada período se puede extender hasta el inicio del siguiente período, o hasta el final de la presentación de medios en el caso del último período. Cada período puede contener una o más representaciones para el mismo contenido de medios. Una

representación puede ser una de varias versiones codificadas alternativas de audio, vídeo, texto temporizado u otros datos de ese tipo. Las representaciones pueden diferir según el tipo de codificación, por ejemplo, según la tasa de bits, la resolución y/o el códec para los datos de vídeo y la tasa de bits, el idioma y/o el códec para los datos de audio. El término representación se puede usar para referirse a una sección de datos de audio o vídeo codificados correspondientes a un período en particular del contenido multimedia y codificados de una forma en particular.

[0017] Las representaciones de un período en particular se pueden asignar a un grupo indicado en la MPD por un atributo indicativo de un conjunto de adaptación al que pertenecen las representaciones. Las representaciones en el mismo conjunto de adaptación en general se consideran alternativas entre sí, en la medida en que un dispositivo cliente puede conmutar entre estas representaciones de manera dinámica y sin interrupciones, por ejemplo, para realizar una adaptación de ancho de banda. Por ejemplo, cada representación de datos de vídeo para un período particular se puede asignar al mismo conjunto de adaptación, de modo que cualquiera de las representaciones se puede seleccionar para descodificar a fin de presentar datos de medios, tales como datos de vídeo o datos de audio, del contenido multimedia para el período correspondiente. El contenido de medios dentro de un período se puede representar mediante una representación cualquiera del grupo 0, si está presente, o bien la combinación de como máximo una representación de cada grupo no cero, en algunos ejemplos. Los datos de temporización para cada representación de un período se pueden expresar con respecto al tiempo de inicio del período.

[0018] Una representación puede incluir uno o más segmentos. Cada representación puede incluir un segmento de inicialización, o cada segmento de una representación puede ser autoinicializador. Cuando está presente, el segmento de inicialización puede contener información de inicialización para acceder a la representación. En general, el segmento de inicialización no contiene datos de medios. Un segmento se puede referenciar de forma única mediante un identificador, tal como un localizador de recursos uniforme (URL), un nombre de recursos uniforme (URN) o un identificador de recursos uniforme (URI). La MPD puede proporcionar los identificadores para cada segmento. En algunos ejemplos, la MPD también puede proporcionar intervalos de bytes en forma de un atributo de *intervalo*, que puede corresponder a los datos para un segmento dentro de un archivo accesible por el URL, el URN o el URI

[0019] Se pueden seleccionar diferentes representaciones para una recuperación sustancialmente simultánea para diferentes tipos de datos de medios. Por ejemplo, un dispositivo cliente puede seleccionar una representación de audio, una representación de vídeo y una representación de texto temporizado a partir de las cuales se pueden recuperar segmentos. En algunos ejemplos, el dispositivo cliente puede seleccionar conjuntos de adaptación particulares para realizar una adaptación de ancho de banda. Es decir, el dispositivo cliente puede seleccionar un conjunto de adaptación que incluye representaciones de vídeo, un conjunto de adaptación que incluye representaciones de audio y/o un conjunto de adaptación que incluye texto temporizado. De forma alternativa, el dispositivo cliente puede seleccionar conjuntos de adaptación para determinados tipos de medios (por ejemplo, vídeo) y seleccionar directamente representaciones para otros tipos de medios (por ejemplo, audio y/o texto temporizado).

[0020] La FIG. 1 es un diagrama de bloques que ilustra un sistema 10 de ejemplo que implementa técnicas para transmitir en continuo datos de medios a través de una red. En este ejemplo, el sistema 10 incluye un dispositivo de preparación de contenido 20, un dispositivo servidor 60 y un dispositivo cliente 40. El dispositivo cliente 40 y el dispositivo servidor 60 están acoplados de forma comunicativa por una red 74, que puede comprender Internet. En algunos ejemplos, el dispositivo de preparación de contenido 20 y el dispositivo servidor 60 también pueden estar acoplados por la red 74 u otra red, o pueden estar directamente acoplados de forma comunicativa. En algunos ejemplos, el dispositivo de preparación de contenido 20 y el dispositivo servidor 60 pueden comprender el mismo dispositivo.

[0021] El dispositivo de preparación de contenido 20, en el ejemplo de la FIG. 1, comprende una fuente de audio 22 y una fuente de vídeo 24. La fuente de audio 22 puede comprender, por ejemplo, un micrófono que produce señales eléctricas representativas de datos de audio captados que el codificador de audio 26 va a codificar. De forma alternativa, la fuente de audio 22 puede comprender un medio de almacenamiento que almacena datos de audio previamente registrados, un generador de datos de audio tal como un sintetizador informatizado, o cualquier otra fuente de datos de audio. La fuente de vídeo 24 puede comprender una cámara de vídeo que produce datos de vídeo que el codificador de vídeo 28 va a codificar, un medio de almacenamiento codificado con datos de vídeo previamente registrados, una unidad de generación de datos de vídeo, tal como una fuente de gráficos de ordenador, o cualquier otra fuente de datos de vídeo. El dispositivo de preparación de contenido 20 no está necesariamente acoplado de forma comunicativa al dispositivo servidor 60 en todos los ejemplos, pero puede almacenar contenido multimedia en un medio separado que el dispositivo servidor 60 lee.

[0022] Los datos de audio y vídeo no procesados pueden comprender datos analógicos o digitales. Los datos analógicos se pueden digitalizar antes de que el codificador de audio 26 y/o el codificador de vídeo 28 los codifiquen. La fuente de audio 22 puede obtener datos de audio a partir de un participante que habla mientras el participante que habla está hablando, y la fuente de vídeo 24 puede obtener simultáneamente datos de vídeo del participante que habla. En otros ejemplos, la fuente de audio 22 puede comprender un medio de almacenamiento legible por ordenador que comprende datos de audio almacenados, y la fuente de vídeo 24 puede comprender un medio de almacenamiento legible por ordenador que comprende datos de vídeo almacenados. De esta manera, las técnicas descritas en esta divulgación se pueden aplicar a la transmisión continua en directo y en tiempo real de datos de audio y vídeo, o de

datos de audio y vídeo archivados y preregistrados.

[0023] Las tramas de audio que corresponden a tramas de vídeo son en general tramas de audio que contienen datos de audio que la fuente de audio 22 ha captado (o generado) al mismo tiempo que unos datos de vídeo, que la fuente de vídeo 24 ha captado (o generado), que están contenidos dentro de las tramas de vídeo. Por ejemplo, mientras un participante que habla en general produce, al hablar, datos de audio, la fuente de audio 22 capta los datos de audio, y la fuente de vídeo 24 capta los datos de vídeo del participante que habla al mismo tiempo, es decir, mientras la fuente de audio 22 está captando los datos de audio. Así pues, una trama de audio puede corresponder temporalmente a una o más tramas de vídeo en particular. Por consiguiente, una trama de audio correspondiente a una trama de vídeo corresponde en general a una situación en la que se han captado datos de audio y datos de vídeo al mismo tiempo, y para la que una trama de audio y una trama de vídeo comprenden, respectivamente, los datos de audio y los datos de vídeo que se han captado al mismo tiempo.

[0024] En algunos ejemplos, el codificador de audio 26 puede codificar una marca de tiempo en cada trama de audio codificada, que representa un tiempo en el que se han registrado los datos de audio para la trama de audio codificada y, de forma similar, el codificador de vídeo 28 puede codificar una marca de tiempo en cada trama de vídeo codificada, que representa un tiempo en el que se han registrado los datos de vídeo para la trama de vídeo codificada. En dichos ejemplos, una trama de audio correspondiente a una trama de vídeo puede comprender una trama de audio que comprende una marca de tiempo y una trama de vídeo que comprende la misma marca de tiempo. El dispositivo de preparación de contenido 20 puede incluir un reloj interno a partir del cual el codificador de audio 26 y/o el codificador de vídeo 28 pueden generar las marcas de tiempo, o que la fuente de audio 22 y la fuente de vídeo 24 pueden usar para asociar datos de audio y vídeo, respectivamente, a una marca de tiempo.

[0025] En algunos ejemplos, la fuente de audio 22 puede enviar datos al codificador de audio 26, correspondientes a un tiempo en el que se han registrado los datos de audio, y la fuente de vídeo 24 puede enviar datos al codificador de vídeo 28, correspondientes a un tiempo en el que se han registrado los datos de vídeo. En algunos ejemplos, el codificador de audio 26 puede codificar un identificador de secuencia en unos datos de audio codificados para indicar un orden temporal relativo de los datos de audio codificados, pero sin indicar necesariamente un tiempo absoluto en el que se han registrado los datos de audio y, de forma similar, el codificador de vídeo 28 también puede usar identificadores de secuencia para indicar un orden temporal relativo de los datos de vídeo codificados. De forma similar, en algunos ejemplos, un identificador de secuencia se puede asignar a, o en cualquier caso correlacionar con, una marca de tiempo.

[0026] El codificador de audio 26, en general, produce un flujo de datos de audio codificados, mientras que el codificador de vídeo 28 produce un flujo de datos de vídeo codificados. Cada flujo de datos individual (ya sea de audio o vídeo) se puede denominar flujo elemental. Un flujo elemental es un componente único codificado digitalmente (y posiblemente comprimido) de una representación. Por ejemplo, la parte de vídeo o audio codificado de la representación puede ser un flujo elemental. Un flujo elemental se puede convertir en un flujo elemental paquetizado (PES) antes de encapsularse dentro de un archivo de vídeo. Dentro de la misma representación, se puede usar un ID de flujo para distinguir los paquetes PES que pertenecen a un flujo elemental de los otros. La unidad básica de datos de un flujo elemental es un paquete de flujo elemental paquetizado (PES). Por tanto, los datos de vídeo codificados corresponden en general a flujos de vídeo elementales. De forma similar, los datos de audio corresponden a uno o más flujos elementales respectivos.

[0027] Muchas normas de codificación de vídeo, tales como ITU-T H.264/AVC y la inminente norma de codificación de vídeo de alta eficacia (HEVC), definen la sintaxis, la semántica y el proceso de descodificación para flujos de bits sin errores, cualquiera de los cuales se ajusta a un determinado perfil o nivel. Las normas de codificación de vídeo típicamente no especifican el codificador, pero el codificador se ocupa de garantizar que los flujos de bits generados cumplan las normas para un descodificador. En el contexto de las normas de codificación de vídeo, un "perfil" corresponde a un subconjunto de algoritmos, rasgos característicos o herramientas y restricciones que se les aplican. Como se define en la norma H.264, por ejemplo, un "perfil" es un subconjunto de toda la sintaxis del flujo de bits especificada por la norma H.264. Un "nivel" corresponde a las limitaciones del consumo de recursos del descodificador, tales como, por ejemplo, memoria y cálculo del descodificador, que están relacionados con la resolución de las imágenes, la velocidad de bits y la velocidad de procesamiento de bloques. Un perfil se puede señalizar con un valor idc de perfil (indicador de perfil), mientras que un nivel se puede señalizar con un valor level_idc (indicador de nivel).

[0028] La norma H.264, por ejemplo, reconoce que, dentro de los límites impuestos por la sintaxis de un perfil dado, todavía es posible requerir una gran variación del rendimiento de los codificadores y descodificadores, dependiendo de los valores adoptados por los elementos de sintaxis en el flujo de bits, tales como el tamaño especificado de las imágenes descodificadas. La norma H.264 reconoce, además, que, en muchas aplicaciones, no es ni práctico ni económico implementar un descodificador capaz de tratar todos los usos hipotéticos de la sintaxis dentro de un perfil en particular. Por consiguiente, la norma H.264 define un "nivel" como un conjunto especificado de restricciones impuestas a los valores de los elementos de sintaxis en el flujo de bits. Estas restricciones pueden ser simples limitaciones de valores. De forma alternativa, estas restricciones pueden adoptar la forma de restricciones sobre combinaciones aritméticas de valores (por ejemplo, la anchura de la imagen multiplicada por la altura de la imagen

multiplicada por el número de imágenes descodificadas por segundo). La norma H.264 establece, además, que las implementaciones individuales pueden admitir un nivel diferente para cada perfil admitido.

[0029] Un descodificador que se ajusta a un perfil normalmente admite todos los rasgos característicos definidos en el perfil. Por ejemplo, como rasgo característico de codificación, la codificación de imágenes B no está admitida en el perfil de línea de base de H.264/AVC, pero está admitida en otros perfiles de H.264/AVC. Un descodificador que se ajusta a un nivel deberá ser capaz de descodificar cualquier flujo de bits que no requiere recursos fuera de las limitaciones definidas en el nivel. Las definiciones de perfiles y niveles pueden ser útiles para la interpretabilidad. Por ejemplo, durante la transmisión de vídeo, se pueden negociar y acordar un par de definiciones de perfil y nivel para una sesión de transmisión completa. Más específicamente, en H.264/AVC, un nivel puede definir limitaciones en el número de macrobloques que es necesario procesar, el tamaño de la memoria intermedia de imágenes descodificadas (DPB), el tamaño de la memoria intermedia de imágenes codificadas (CPB), el intervalo de vectores de movimiento vertical, el número máximo de vectores de movimiento para cada dos MB consecutivos y si un bloque B puede tener divisiones de submacrobloque inferiores a 8x8 píxeles. De esta manera, un descodificador puede determinar si el descodificador es capaz de descodificar apropiadamente el flujo de bits.

[0030] En el ejemplo de la FIG. 1, la unidad de encapsulación 30 del dispositivo de preparación de contenido 20 recibe flujos elementales que comprenden datos de vídeo codificados desde el codificador de vídeo 28 y flujos elementales que comprenden datos de audio codificados desde el codificador de audio 26. En algunos ejemplos, el codificador de vídeo 28 y el codificador de audio 26 pueden incluir, cada uno, paquetizadores para formar paquetes PES a partir de datos codificados. En otros ejemplos, el codificador de vídeo 28 y el codificador de audio 26 pueden interactuar, cada uno, con los paquetizadores respectivos para formar paquetes PES a partir de datos codificados. En otros ejemplos más, la unidad de encapsulación 30 puede incluir paquetizadores para formar paquetes PES a partir de datos de audio y de vídeo codificados.

[0031] El codificador de vídeo 28 puede codificar datos de vídeo de contenido multimedia en una variedad de formas, para producir diferentes representaciones del contenido multimedia a diversas tasas de bits y con diversas características, tales como resoluciones de píxeles, velocidades de tramas, conformidad con diversas normas de codificación, conformidad con diversos perfiles y/o niveles de perfiles para diversas normas de codificación, representaciones que tienen una o múltiples vistas (por ejemplo, para reproducción bidimensional o tridimensional), u otras características de ese tipo. Una representación, como se usa en esta divulgación, puede comprender uno de datos de audio, datos de vídeo, datos de texto (por ejemplo, para subtítulos cerrados) u otros datos de este tipo. La representación puede incluir un flujo elemental, tal como un flujo elemental de audio o un flujo elemental de vídeo. Cada paquete PES puede incluir un identificador stream_id que identifica el flujo elemental al que pertenece el paquete PES. La unidad de encapsulación 30 es responsable de ensamblar flujos elementales en archivos de vídeo (por ejemplo, segmentos) de diversas representaciones.

[0032] La unidad de encapsulación 30 recibe paquetes PES para flujos elementales de una representación desde el codificador de audio 26 y el codificador de vídeo 28 y forma las correspondientes unidades de capa de abstracción de red (NAL) a partir de los paquetes PES. En el ejemplo de la H.264/AVC (codificación de vídeo avanzada), los segmentos de vídeo codificados están organizados en unidades de NAL, que proporcionan una representación de vídeo "apta para redes" dirigida a aplicaciones tales como la videotelefonia, el almacenamiento, la radiodifusión o la transmisión continua. Las unidades NAL se pueden clasificar en unidades NAL de capa de codificación de vídeo (VCL) y unidades NAL no VCL. Las unidades VCL pueden contener el motor de compresión central y pueden incluir datos a nivel de bloque, macrobloque y/o fragmento. Otras unidades NAL pueden ser unidades NAL no VCL. En algunos ejemplos, una imagen codificada en una instancia de tiempo, normalmente presentada como una imagen codificada primaria, puede estar contenida en una unidad de acceso, que puede incluir una o más unidades NAL.

[0033] Las unidades NAL no VCL pueden incluir unidades NAL de conjunto de parámetros y unidades NAL SEI, entre otras. Los conjuntos de parámetros pueden contener información de cabecera a nivel de secuencia (en conjuntos de parámetros de secuencia (SPS)) y la información de cabecera a nivel de imagen que cambia ocasionalmente (en conjuntos de parámetros de imagen (PPS)). Con los conjuntos de parámetros (por ejemplo, PPS y SPS), la información que cambia ocasionalmente no necesita ser repetida para cada secuencia o imagen, de ahí que pueda mejorarse la eficacia de la codificación. Además, el uso de conjuntos de parámetros puede permitir la transmisión fuera de banda de la información de cabecera importante, evitando la necesidad de transmisiones redundantes para la resistencia a los errores. En los ejemplos de transmisión fuera de banda, las unidades NAL de conjunto de parámetros se pueden transmitir en un canal diferente al de otras unidades NAL, tales como las unidades NAL SEI.

[0034] La información de mejora complementaria (SEI) puede contener información que no es necesaria para descodificar las muestras de imágenes codificadas a partir de las unidades NAL VCL, pero puede ayudar en los procesos relacionados con la descodificación, visualización, resistencia a los errores y otros propósitos. Los mensajes de SEI pueden estar contenidos en unidades NAL no VCL. Los mensajes de SEI son la parte normativa de algunas memorias descriptivas habituales y, por tanto, no siempre son obligatorios para la implementación de descodificadores que cumplen las normas. Los mensajes de SEI pueden ser mensajes de SEI a nivel de secuencia o mensajes de SEI a nivel de imagen. Parte de la información a nivel de secuencia puede estar contenida en mensajes de SEI, tales como mensajes de SEI de información de adaptabilidad a escala en el ejemplo de SVC y mensajes de SEI de

información de adaptabilidad a escala de la vista en MVC. Estos ejemplos de mensajes de SEI pueden transmitir información, por ejemplo, sobre extracción de puntos de funcionamiento y características de los puntos de funcionamiento. Además, la unidad de encapsulación 30 puede formar un archivo de manifiesto, tal como una descripción de presentación de medios (MPD) que describe características de las representaciones. La unidad de encapsulación 30 puede formatear la MPD de acuerdo con un lenguaje de marcado extensible (XML).

[0035] La unidad de encapsulación 30 puede proporcionar datos para una o más representaciones de contenido multimedia, junto con el archivo de manifiesto (por ejemplo, la MPD), a la interfaz de salida 32. La interfaz de salida 32 puede comprender una interfaz de red o una interfaz para escribir en un medio de almacenamiento, tal como una interfaz de bus serie universal (USB), una grabadora o copiadora de CD o DVD, una interfaz para medios de almacenamiento magnéticos o flash, u otras interfaces para almacenar o transmitir datos de medios. La unidad de encapsulación 30 puede proporcionar datos de cada una de las representaciones de contenido multimedia a la interfaz de salida 32, que puede enviar los datos al dispositivo servidor 60 por medio de transmisión por red o medios de almacenamiento. En el ejemplo de la FIG. 1, el dispositivo servidor 60 incluye un medio de almacenamiento 62 que almacena diversos contenidos multimedia 64, incluyendo cada uno un respectivo archivo de manifiesto 66 y una o más representaciones 68A a 68N (representaciones 68). En algunos ejemplos, la interfaz de salida 32 también puede enviar datos directamente a la red 74.

[0036] En algunos ejemplos, las representaciones 68 se pueden separar en conjuntos de adaptación. Es decir, diversos subconjuntos de representaciones 68 pueden incluir respectivos conjuntos comunes de características, tales como códec, perfil y nivel, resolución, número de vistas, formato de archivo para segmentos, información de tipo de texto que puede identificar un idioma u otras características de un texto que se va a visualizar con la representación y/o datos de audio que se van a descodificar y presentar, por ejemplo, mediante altavoces, información de ángulo de cámara que puede describir un ángulo de cámara o una perspectiva de cámara real de una escena para representaciones del conjunto de adaptación, información de calificación que describe la idoneidad del contenido para audiencias en particular, o similares.

[0037] El archivo de manifiesto 66 puede incluir datos indicativos de los subconjuntos de representaciones 68 correspondientes a conjuntos de adaptación en particular, así como características comunes para los conjuntos de adaptación. El archivo de manifiesto 66 también puede incluir datos representativos de características individuales, tales como las tasas de bits, para representaciones individuales de conjuntos de adaptación. De esta manera, un conjunto de adaptación puede hacer posible una adaptación simplificada del ancho de banda de red. Las representaciones de un conjunto de adaptación se pueden indicar usando elementos hijo de un elemento del conjunto de adaptación del archivo de manifiesto 66.

[0038] El dispositivo servidor 60 incluye una unidad de procesamiento de peticiones 70 y una interfaz de red 72. En algunos ejemplos, el dispositivo servidor 60 puede incluir una pluralidad de interfaces de red. Además, uno cualquiera o todos los rasgos característicos del dispositivo servidor 60 se pueden implementar en otros dispositivos de una red de entrega de contenido, tales como encaminadores, puentes, dispositivos proxy, conmutadores u otros dispositivos. En algunos ejemplos, los dispositivos intermedios de una red de entrega de contenido pueden almacenar en memoria caché datos de contenido multimedia 64, e incluir componentes que se ajustan sustancialmente a los del dispositivo servidor 60. En general, la interfaz de red 72 está configurada para enviar y recibir datos por medio de la red 74.

[0039] La unidad de procesamiento de peticiones 70 está configurada para recibir peticiones de red desde dispositivos cliente, tales como el dispositivo cliente 40, para datos del medio de almacenamiento 62. Por ejemplo, la unidad de procesamiento de peticiones 70 puede implementar el protocolo de transferencia de hipertexto (HTTP) versión 1.1, como se describe en RFC 2616, "Hypertext Transfer Protocol - HTTP/1.1", de R. Fielding y otros, Grupo de Trabajo de la Red, IETF, junio de 1999. Es decir, la unidad de procesamiento de peticiones 70 puede estar configurada para recibir peticiones GET o GET parciales HTTP y proporcionar datos de contenido multimedia 64 como respuesta a las peticiones. Las peticiones pueden especificar un segmento de una de las representaciones 68, por ejemplo, usando un URL del segmento. En algunos ejemplos, las peticiones también pueden especificar uno o más intervalos de bytes del segmento, comprendiendo por tanto peticiones GET parciales. La unidad de procesamiento de peticiones 70 puede estar configurada, además, para atender peticiones HEAD HTTP para proporcionar datos de cabecera de un segmento de una de las representaciones 68. En cualquier caso, la unidad de procesamiento de peticiones 70 puede estar configurada para procesar las peticiones para proporcionar los datos solicitados a un dispositivo solicitante, tal como el dispositivo cliente 40.

[0040] De forma adicional o alternativa, la unidad de procesamiento de peticiones 70 puede estar configurada para entregar datos de medios por medio de un protocolo de radiodifusión o multidifusión, tal como el eMBMS. El dispositivo de preparación de contenido 20 puede crear segmentos y/o subsegmentos DASH, sustancialmente de la misma manera que se ha descrito, pero el dispositivo servidor 60 puede entregar estos segmentos o subsegmentos usando el eMBMS u otro protocolo de transporte de red de radiodifusión o multidifusión. Por ejemplo, la unidad de procesamiento de peticiones 70 puede estar configurada para recibir una petición para unirse a un grupo de multidifusión desde el dispositivo cliente 40. Es decir, el dispositivo servidor 60 puede comunicar una dirección de protocolo de Internet (IP), asociada a un grupo de multidifusión a unos dispositivos cliente, que incluyen el dispositivo cliente 40, asociados a un contenido de medios en particular (por ejemplo, radiodifusión de un acontecimiento en

directo). El dispositivo cliente 40, a su vez, puede presentar una petición para unirse al grupo de multidifusión. Esta petición se puede propagar por toda la red 74, por ejemplo, los encaminadores que componen la red 74, de modo que se hace que los encaminadores dirijan el tráfico destinado a la dirección IP asociada al grupo de multidifusión a los dispositivos cliente abonados, tales como el dispositivo cliente 40.

[0041] Como se ilustra en el ejemplo de la FIG. 1, el contenido multimedia 64 incluye el archivo de manifiesto 66, que puede corresponder a una descripción de presentación de medios (MPD). El archivo de manifiesto 66 puede contener descripciones de diferentes representaciones alternativas 68 (por ejemplo, servicios de vídeo con diferentes calidades) y la descripción puede incluir, por ejemplo, información de códec, un valor de perfil, un valor de nivel, una tasa de bits y otras características descriptivas de las representaciones 68. El dispositivo cliente 40 puede recuperar la MPD de una presentación de medios para determinar cómo acceder a segmentos de las representaciones 68.

[0042] En particular, la unidad de recuperación 52 puede recuperar datos de configuración (no mostrados) del dispositivo cliente 40 para determinar las capacidades de descodificación del descodificador de vídeo 48 y las capacidades de representación de la salida de vídeo 44. Los datos de configuración también pueden incluir cualquiera o todas las preferencias de idioma seleccionadas por un usuario del dispositivo cliente 40, una o más perspectivas de cámara correspondientes a las preferencias de profundidad establecidas por el usuario del dispositivo cliente 40 y/o una preferencia de calificación seleccionada por el usuario del dispositivo cliente 40. La unidad de recuperación 52 puede comprender, por ejemplo, un navegador web o un cliente de medios configurados para presentar peticiones GET y GET parcial HTTP. La unidad de recuperación 52 puede corresponder a unas instrucciones de software ejecutadas por uno o más procesadores o unidades de procesamiento (no mostrados) del dispositivo cliente 40. En algunos ejemplos, la totalidad o unas partes de la funcionalidad descrita con respecto a la unidad de recuperación 52 se pueden implementar en hardware, o una combinación de hardware, software y/o firmware, donde se puede proporcionar el hardware requerido para ejecutar instrucciones para software o firmware.

[0043] La unidad de recuperación 52 puede comparar las capacidades de descodificación y representación del dispositivo cliente 40 con las características de las representaciones 68 indicadas por la información del archivo de manifiesto 66. La unidad de recuperación 52 puede recuperar inicialmente al menos una parte del archivo de manifiesto 66 para determinar las características de las representaciones 68. Por ejemplo, la unidad de recuperación 52 puede solicitar una parte del archivo de manifiesto 66 que describe las características de uno o más conjuntos de adaptación. La unidad de recuperación 52 puede seleccionar un subconjunto de representaciones 68 (por ejemplo, un conjunto de adaptación) que tiene características que se pueden satisfacer mediante las capacidades de codificación y representación del dispositivo cliente 40. La unidad de recuperación 52 puede, a continuación, determinar las tasas de bits para las representaciones del conjunto de adaptación, determinar una cantidad actualmente disponible de ancho de banda de red y recuperar segmentos de una de las representaciones que tiene una tasa de bits que se puede satisfacer mediante el ancho de banda de red.

[0044] En general, las representaciones de tasas de bits mayores pueden producir una reproducción de vídeo de mayor calidad, mientras que las representaciones de tasas de bits más bajas pueden proporcionar una reproducción de vídeo de calidad suficiente cuando el ancho de banda de red disponible se reduce. En consecuencia, cuando el ancho de banda de red disponible es relativamente alto, la unidad de recuperación 52 puede recuperar datos de representaciones de tasas de bits relativamente altas, mientras que cuando el ancho de banda de red disponible es bajo, la unidad de recuperación 52 puede recuperar datos de representaciones de tasas de bits relativamente bajas. De esta manera, el dispositivo cliente 40 puede transmitir en continuo datos de medios a través de la red 74 mientras que también se adapta a la disponibilidad cambiante de ancho de banda de red de la red 74.

[0045] De forma adicional o alternativa, la unidad de recuperación 52 puede estar configurada para recibir datos de acuerdo con un protocolo de red de radiodifusión o multidifusión, tal como la multidifusión eMBMS o IP. En dichos ejemplos, la unidad de recuperación 52 puede presentar una petición para unirse a un grupo de red de multidifusión asociado a un contenido de medios en particular. Después de unirse al grupo de multidifusión, la unidad de recuperación 52 puede recibir datos del grupo de multidifusión sin peticiones adicionales emitidas al dispositivo servidor 60 o al dispositivo de preparación de contenido 20. La unidad de recuperación 52 puede presentar una petición para abandonar el grupo de multidifusión cuando ya no se necesitan datos del grupo de multidifusión, por ejemplo, para detener la reproducción o para cambiar canales a un grupo de multidifusión diferente.

[0046] La interfaz de red 54 puede recibir y proporcionar datos de segmentos de una representación seleccionada a la unidad de recuperación 52, que a su vez puede proporcionar los segmentos a la unidad de procesamiento del formato de archivo 50. La unidad de procesamiento del formato de archivo 50 puede desencapsular elementos de un archivo de vídeo en flujos PES constituyentes, desempaquetar los flujos PES para recuperar datos codificados y enviar los datos codificados al descodificador de audio 46 o bien al descodificador de vídeo 48, dependiendo de si los datos codificados forman parte de un flujo de audio o vídeo, por ejemplo, como lo indican las cabeceras de paquetes PES del flujo. El descodificador de audio 46 descodifica datos de audio codificados y envía los datos de audio descodificados a la salida de audio 42, mientras que el descodificador de vídeo 48 descodifica datos de vídeo codificados y envía los datos de vídeo descodificados, que pueden incluir una pluralidad de vistas de un flujo, a la salida de vídeo 44.

[0047] El codificador de vídeo 28, el descodificador de vídeo 48, el codificador de audio 26, el descodificador de audio 46, la unidad de encapsulación 30, la unidad de recuperación 52 y la unidad de procesamiento del formato de archivo 50 se pueden implementar cada uno como cualquiera de una variedad de circuitos de procesamiento adecuados, según corresponda, tales como uno o más microprocesadores, procesadores de señales digitales (DSP), circuitos integrados específicos de la aplicación (ASIC), matrices de puertas programables in situ (FPGA), circuitos lógicos discretos, software, hardware, firmware o cualquier combinación de los mismos. Tanto el codificador de vídeo 28 como el descodificador de vídeo 48 pueden estar incluidos en uno o más codificadores o descodificadores, ambos de los cuales pueden estar integrados como parte de un codificador/descodificador (CÓDEC) de vídeo combinado. Asimismo, tanto el codificador de audio 26 como el descodificador de audio 46 pueden estar incluidos en uno o más codificadores o descodificadores, ambos de los cuales pueden estar integrados como parte de un CÓDEC combinado. Un aparato que incluye un codificador de vídeo 28, un descodificador de vídeo 48, un codificador de audio 26, un descodificador de audio 46, una unidad de encapsulación 30, una unidad de recuperación 52 y/o una unidad de procesamiento del formato de archivo 50 puede comprender un circuito integrado, un microprocesador y/o un dispositivo de comunicación inalámbrica, tal como un teléfono móvil.

[0048] El dispositivo cliente 40, el dispositivo servidor 60 y/o el dispositivo de preparación de contenido 20 pueden estar configurados para funcionar de acuerdo con las técnicas de esta divulgación. Con propósitos de ejemplo, esta divulgación describe estas técnicas con respecto al dispositivo cliente 40 y al dispositivo servidor 60. Sin embargo, se deberá entender que el dispositivo de preparación de contenido 20 puede estar configurado para realizar estas técnicas, en lugar (o, además) del dispositivo servidor 60.

[0049] La unidad de encapsulación 30 puede formar unidades NAL que comprenden una cabecera que identifica un programa al cual pertenece la unidad NAL, así como una carga útil, por ejemplo, datos de audio, datos de vídeo o datos que describen el flujo de transporte o de programa al cual corresponde la unidad NAL. Por ejemplo, en H.264/AVC, una unidad NAL incluye una cabecera de 1 byte y una carga útil de tamaño variable. Una unidad NAL que incluye datos de vídeo en su carga útil puede comprender diversos niveles de granularidad de datos de vídeo. Por ejemplo, una unidad NAL puede comprender un bloque de datos de vídeo, una pluralidad de bloques, un fragmento de datos de vídeo o una imagen completa de datos de vídeo. La unidad de encapsulación 30 puede recibir datos de vídeo codificados desde el codificador de vídeo 28 en forma de paquetes PES de flujos elementales. La unidad de encapsulación 30 puede asociar cada flujo elemental a un programa correspondiente.

[0050] La unidad de encapsulación 30 también puede ensamblar unidades de acceso desde una pluralidad de unidades NAL. En general, una unidad de acceso puede comprender una o más unidades NAL para representar una trama de datos de vídeo, así como datos de audio correspondientes a la trama cuando dichos datos de audio están disponibles. Una unidad de acceso incluye en general todas las unidades NAL para una instancia de tiempo de salida, por ejemplo, todos los datos de audio y vídeo para una instancia de tiempo. Por ejemplo, si cada visualización tiene una frecuencia de tramas de 20 tramas por segundo (fps), cada instancia de tiempo puede corresponder a un intervalo de tiempo de 0,05 segundos. Durante este intervalo de tiempo, las tramas específicas para todas las vistas de la misma unidad de acceso (la misma instancia de tiempo) se pueden representar simultáneamente. En un ejemplo, una unidad de acceso puede comprender una imagen codificada en una instancia de tiempo, que se puede presentar como una imagen codificada primaria.

[0051] Por consiguiente, una unidad de acceso puede comprender todas las tramas de audio y vídeo de una instancia temporal común, por ejemplo, todas las vistas correspondientes al tiempo X. Esta divulgación también se refiere a una imagen codificada de una vista particular como un "componente de vista". Es decir, un componente de vista puede comprender una imagen (o trama) codificada para una vista en particular en un tiempo en particular. Por consiguiente, se puede definir una unidad de acceso que comprende todos los componentes de vista de una instancia temporal común. El orden de descodificación de las unidades de acceso no necesita ser el mismo que el orden de salida o de visualización.

[0052] Una presentación de medios puede incluir una descripción de presentación de medios (MPD), que puede contener descripciones de diferentes representaciones alternativas (por ejemplo, servicios de vídeo con diferentes calidades) y la descripción puede incluir, por ejemplo, información de códec, un valor de perfil y un valor de nivel. Una MPD es un ejemplo de archivo de manifiesto, tal como el archivo de manifiesto 66. El dispositivo cliente 40 puede recuperar la MPD de una presentación de medios para determinar cómo acceder a fragmentos de película de diversas presentaciones. Los fragmentos de película pueden estar situados en cuadros de fragmento de película (cuadros moof) de archivos de vídeo.

[0053] El archivo de manifiesto 66 (que puede comprender, por ejemplo, una MPD) puede comunicar la disponibilidad de segmentos de representaciones 68. Es decir, la MPD puede incluir información que indica el tiempo de hora de reloj en el cual un primer segmento de una de las representaciones 68 queda disponible, así como información que indica las duraciones de los segmentos dentro de las representaciones 68. De esta manera, la unidad de recuperación 52 del dispositivo cliente 40 puede determinar cuándo está disponible cada segmento, en base al tiempo de inicio, así como de las duraciones de los segmentos que preceden a un segmento en particular.

[0054] Después de que la unidad de encapsulación 30 haya ensamblado las unidades de NAL y/o las unidades de

acceso en un archivo de vídeo, en base a los datos recibidos, la unidad de encapsulación 30 pasa el archivo de vídeo a la interfaz de salida 32 para su salida. En algunos ejemplos, la unidad de encapsulación 30 puede almacenar el archivo de vídeo localmente o enviar el archivo de vídeo a un servidor remoto por medio de la interfaz de salida 32, en lugar de enviar el archivo de vídeo directamente al dispositivo cliente 40. La interfaz de salida 32 puede comprender, por ejemplo, un transmisor, un transceptor, un dispositivo para escribir datos en un medio legible por ordenador tal como, por ejemplo, una unidad óptica, una unidad de medios magnéticos (por ejemplo, una unidad de disquetes), un puerto de bus serie universal (USB), una interfaz de red u otra interfaz de salida. La interfaz de salida 32 emite el archivo de vídeo a un medio legible por ordenador, tal como, por ejemplo, una señal de transmisión, un medio magnético, un medio óptico, una memoria, una unidad flash u otro medio legible por ordenador.

[0055] La interfaz de red 54 puede recibir una unidad NAL o unidad de acceso por medio de la red 74 y proporcionar la unidad NAL o la unidad de acceso a la unidad de procesamiento del formato de archivo 50, por medio de la unidad de recuperación 52. La unidad de procesamiento del formato de archivo 50 puede desencapsular un elemento de un archivo de vídeo en flujos PES constituyentes, desempaquetar los flujos PES para recuperar los datos codificados y enviar los datos codificados al descodificador de audio 46 o al descodificador de vídeo 48, dependiendo de si los datos codificados forman parte de un flujo de audio o vídeo, por ejemplo, como se indica en las cabeceras de paquetes PES del flujo. El descodificador de audio 46 descodifica datos de audio codificados y envía los datos de audio descodificados a la salida de audio 42, mientras que el descodificador de vídeo 48 descodifica datos de vídeo codificados y envía los datos de vídeo descodificados, que pueden incluir una pluralidad de vistas de un flujo, a la salida de vídeo 44.

[0056] El dispositivo de preparación de contenido 20 (a través de, por ejemplo, la unidad de encapsulación 30) y el dispositivo cliente 40 (a través de, por ejemplo, la unidad de procesamiento del formato de archivo 50) pueden utilizar uno o varios de muchos formatos de archivo para encapsular y/o desencapsular contenido de vídeo. El ISOBMFF se usa como la base para muchos formatos de encapsulación de códec, tales como el formato de archivo AVC, así como para muchos formatos de contenedor multimedia, tales como el formato de archivo MPEG-4, el formato de archivo 3GPP (3GP) y el formato de archivo DVB.

[0057] Además de los medios continuos, tales como el audio y el vídeo, los medios estáticos, tales como las imágenes, así como los metadatos, se pueden almacenar en un archivo que se ajusta al ISOBMFF. Los archivos estructurados de acuerdo con el ISOBMFF se pueden usar para muchos propósitos, incluyendo la reproducción local de archivos de medios, la descarga progresiva de un archivo remoto, segmentos para la transmisión en continuo dinámica adaptativa a través de HTTP (DASH), contenedores para contenido que se va a transmitir en continuo y sus instrucciones de empaquetado y el registro de flujos de medios recibidos en tiempo real.

[0058] Una caja es la estructura sintáctica elemental en el ISOBMFF, que incluye un tipo de caja codificada de cuatro caracteres, un recuento de bytes de la caja y la carga útil. Un archivo ISOBMFF consiste en una secuencia de cajas, y las cajas pueden contener otras cajas. Una caja de película ("moov") contiene los metadatos para los flujos continuos de medios presentes en el archivo, cada uno representado en el archivo como una pista. Los metadatos para una pista están contenidos en una caja de pista ("trak"), mientras que el contenido de medios de una pista está contenido en una caja de datos de medios ("mdat") o directamente en un archivo separado. El contenido de medios para pistas consiste en una secuencia de muestras, tales como unas unidades de acceso de audio o vídeo.

[0059] El ISOBMFF especifica los siguientes tipos de pistas: una pista de medios, que contiene un flujo de medios elemental, una pista de indicaciones, que incluye instrucciones de transmisión de medios o representa un flujo de paquetes recibido, y una pista de metadatos temporizados, que comprende metadatos sincronizados en el tiempo.

[0060] Aunque originalmente se diseñó para el almacenamiento, el ISOBMFF ha demostrado ser muy valioso para la transmisión en continuo, por ejemplo, para una descarga progresiva o DASH. Para propósitos de transmisión en continuo, se pueden usar los fragmentos de película definidos en el ISOBMFF.

[0061] Los metadatos para cada pista incluyen una lista de entradas de descripción de muestra, cada una de las cuales proporciona el formato de codificación o encapsulación usado en la pista y los datos de inicialización necesarios para procesar ese formato. Cada muestra está asociada a una de las entradas de descripción de muestra de la pista.

[0062] El ISOBMFF permite especificar metadatos específicos de muestra con diversos mecanismos. Las cajas específicas dentro de la caja de tabla de muestras ("stbl") se han normalizado para responder a necesidades comunes. Por ejemplo, se usa una caja de muestras de sincronización ("stss") para enumerar las muestras de acceso aleatorio de la pista. El mecanismo de agrupación de muestras permite la correlación de muestras de acuerdo con un tipo de agrupación de cuatro caracteres en grupos de muestras que comparten la misma propiedad especificada como una entrada de descripción de grupo de muestras en el archivo. Se han especificado diversos tipos de agrupación en el ISOBMFF.

[0063] La especificación ISOBMFF especifica seis tipos de puntos de acceso de flujo (SAP) para usar con DASH. Los dos primeros tipos de SAP (tipos 1 y 2) corresponden a imágenes IDR en H.264/AVC y HEVC. El tercer tipo de SAP (tipo 3) corresponde a puntos de acceso aleatorio de GOP abierto, por consiguiente, a imágenes BLA o CRA en

HEVC. El cuarto tipo de SAP (tipo 4) corresponde a puntos de acceso aleatorio de GDR.

[0064] En ISO/IEC 14496-15, se especifican varios tipos de entrada de muestra (también denominados nombres de entrada de muestra).

[0065] En el formato de archivo HEVC (cláusula 8 de ISO/IEC 14496-15), se especifican los tipos de entrada de muestra 'hvc1' y 'hevl'. Una restricción para el tipo de entrada de muestra 'hevl' se especifica de la siguiente manera:

Cuando el nombre de entrada de muestra es 'hev1', se aplica lo siguiente: Si la muestra es un punto de acceso aleatorio, todos los conjuntos de parámetros necesarios para descodificar esa muestra se incluirán en la entrada de la muestra o en la muestra misma.

[0066] De lo contrario (la muestra no es un punto de acceso aleatorio), todos los conjuntos de parámetros necesarios para descodificar la muestra se incluirán en la entrada de la muestra o en cualquiera de las muestras, desde el punto de acceso aleatorio anterior a esa propia muestra, inclusive.

[0067] El propósito de esta restricción también es permitir un acceso aleatorio conveniente, desde la muestra que es un punto de acceso aleatorio sin la necesidad de buscar y recuperar conjuntos de parámetros de muestras anteriores.

[0068] En el formato de archivo HEVC en capas (L-HEVC) (cláusula 9 de ISO/IEC 14496-15), se especifican los tipos de entrada de muestra 'hvc2', 'hev2', 'lhv1' y 'lhe1'. Una restricción para el tipo de entrada de muestra 'lhe1' se especifica de la siguiente manera:

[0069] Cuando el nombre de entrada de muestra es 'lhe1', se aplica lo siguiente:

Las restricciones a continuación imponen restricciones en la colocación de conjuntos de parámetros fuera de banda (en entradas de muestra) y conjuntos de parámetros dentro de banda (en muestras), para permitir un acceso aleatorio conveniente desde unidades de acceso que contienen imágenes IRAP al menos en algunas capas. Con estas restricciones, un lector de archivos que se inicializa con las entradas de muestra y avanza desde una unidad de acceso en la que todas las imágenes son imágenes IRAP tendrá todos los conjuntos de parámetros que necesita.

[0070] Para cualquier muestra particular en una pista particular, la muestra colocada temporalmente en otra pista se define como aquella con el mismo tiempo de descodificación que la de esta muestra particular.

[0071] Para una imagen IRAP de una muestra, pista y capa dados, cada conjunto de parámetros necesarios para descodificar la imagen IRAP se incluirá en uno de los siguientes:

a. la entrada de muestra que se aplica a la muestra dada en la pista dada

b. la entrada de muestra de la muestra inicial de una pista que lleva una capa de referencia de la capa dada, donde la muestra inicial es la muestra colocada temporalmente de la muestra dada, cuando la muestra colocada temporalmente contiene una imagen IRAP de la capa de referencia, o la muestra anterior que contiene una imagen IRAP de la capa de referencia

c. la muestra dada en sí, posiblemente mediante el uso de extractores

d. cuando está presente, cualquier muestra colocada temporalmente de las pistas que llevan capas de referencia de la capa dada, posiblemente mediante el uso de extractores

[0072] Para una imagen no IRAP de una muestra, pista y capa dados, cada conjunto de parámetros necesarios para descodificar esa imagen se incluirá en uno de los siguientes:

a. la entrada de muestra que se aplica a la muestra dada en la pista dada

b. la entrada de muestra de la muestra inicial de una pista que lleva una capa de referencia de la capa dada, donde la muestra inicial es la muestra colocada temporalmente de la muestra dada, cuando la muestra colocada temporalmente contiene una imagen IRAP de la capa de referencia, o la muestra anterior que contiene una imagen IRAP de la capa de referencia

c. cualquiera de las muestras en la pista dada desde la muestra anterior que contiene una imagen IRAP en la capa dada a la muestra dada en sí, inclusive, posiblemente mediante el uso de extractores

d. cuando está presente, cualquiera de las muestras en una pista que lleva una capa de referencia de la capa dada desde la muestra colocada temporalmente de la muestra anterior que contiene una imagen IRAP en la capa dada hasta la muestra colocada temporalmente de la muestra dada, inclusive, posiblemente usando

extractores

[0073] El propósito de esta restricción también es permitir un acceso aleatorio conveniente, pero para flujos de bits que contienen múltiples capas, como se describe en detalle anteriormente como parte de la descripción de la restricción.

La Tabla 10 de ISO/IEC 14496-15 (copiada a continuación) muestra todos los usos posibles de entradas de muestra, configuraciones y las herramientas L-HEV C para pistas HEVC y L-HEVC:

Nombre de entrada de muestra	**Con registros de configuración**	**Significado**
'hvc1' o 'hev1'	Solo configuración HEVC	Una pista HEVC simple sin unidades NAL con nuh_layer_id mayor que 0; no estarán presentes ni extractores ni agregadores.
'hvc1' o 'hev1'	Configuraciones HEVC y L-HEVC	Una pista L-HEVC con unidades NAL con nuh_layer_id igual a 0 y unidades NAL con nuh_layer_id mayor que 0; Los extractores y agregadores no deben estar presentes.
'hvc2' o 'hev2'	Solo configuración HEVC	Una pista HEVC simple sin unidades NAL con nuh_layer_id mayor que 0; pueden estar presentes y usarse extractores para hacer referencia a unidades NAL; pueden estar presentes agregadores para contener y hacer referencia a unidades NAL.
'hvc2' o 'hev2'	Configuraciones HEVC y L-HEVC	Una pista L-HEVC con unidades NAL con nuh_layer_id igual a 0 y unidades NAL con nuh_layer_id mayor que 0; pueden estar presentes extractores y agregadores; los extractores pueden hacer referencia a cualquier unidad NAL; los agregadores pueden contener y hacer referencia a cualquier unidad NAL.
'lhv1', 'lhe1'	Solo configuración L-HEVC	Una pista L-HEVC con unidades NAL con nuh_layer_id mayor que 0 y sin unidades NAL con nuh_layer_id igual a 0; pueden estar presentes extractores y usarse para hacer referencia a unidades NAL; pueden estar presentes agregadores para contener y hacer referencia a unidades NAL.

[0074] Esta divulgación reconoce que los diseños actuales de tipos de entrada de muestra en los formatos de archivo para HEVC y L-HEVC (por ejemplo, en las cláusulas 8 y 9 de ISO/IEC 14496-15) pueden presentar varios problemas. Por ejemplo:

[0075] Para describir un primer problema potencial, se observa que la fila 2 de la Tabla 10 de ISO/IEC 14496-15 establece:

'hvc1' o 'hev1'	Configuraciones HEVC y L-HEVC	Una pista L-HEVC con unidades NAL con nuh_layer_id igual a 0 y unidades NAL con nuh_layer_id mayor que 0; Los extractores y agregadores no deben estar presentes.

[0076] Mientras tanto, la fila 4 de la Tabla 10 de ISO/IEC 14496-15 establece:

'hvc2' o 'hev2'	Configuraciones HEVC y L-HEVC	Una pista L-HEVC con unidades NAL con nuh_layer_id igual a 0 y unidades NAL con nuh_layer_id mayor que 0; pueden estar presentes extractores y agregadores; los extractores pueden hacer referencia a cualquier unidad NAL; los agregadores pueden contener y hacer referencia a cualquier unidad NAL.

[0077] Cuando el tipo de entrada de muestra es 'hvc1', 'hev1', 'hvc2' o 'hev2', y la misma entrada contiene configuraciones HEVC y LHEVC, la pista transporta datos tanto de la capa base como de una o más capas de mejora.

[0078] Sin embargo, la restricción que permite un acceso aleatorio conveniente pero, para flujos de bits que

contienen múltiples capas, solo se especifica para el tipo de entrada de muestra 'lhe1'. Esto significa que cuando se almacena un flujo de bits L-HEVC multicapa utilizando cualquiera de los tipos de entrada de muestra 'hvc1', 'hev1', 'hvc2' y 'hev2', no hay forma de garantizar e indicar que ese acceso aleatorio conveniente (sin necesidad de buscar y recuperar conjuntos de parámetros de muestras anteriores, etc.), está habilitado.

[0079] Las técnicas de esta divulgación pueden usarse para abordar el primer problema potencial analizado anteriormente. En particular, en un ejemplo, se especifica una restricción para permitir un acceso aleatorio conveniente, sin necesidad de buscar y recuperar conjuntos de parámetros de muestras anteriores para una pista que contiene datos de múltiples capas, para los tipos de entrada de muestra 'hev1' y 'hev2' cuando la entrada de muestra contiene ambas configuraciones HEVC y L-HEVC. Por lo tanto, el dispositivo de preparación de contenido 20 puede garantizar que los conjuntos de parámetros necesarios se proporcionen con entradas de muestra y/o muestras que tengan tipos de entrada de muestra de 'hevl' y 'hev2', de modo que no se requiera buscar ni recuperar conjuntos de parámetros de muestras anteriores. Del mismo modo, el dispositivo cliente 40 puede recuperar una entrada de muestra y una muestra que tiene un tipo de entrada de muestra de 'hevl' o 'hev2' y realizar un acceso aleatorio, sin recuperar conjuntos de parámetros de ninguna muestra previa en el orden de descodificación de vídeo. La Tabla 10 de ISO/IEC 14496-15 (copiada a continuación) muestra todos los usos posibles de entradas de muestra, configuraciones y las herramientas L-HEV C para pistas HEVC y L-HEVC:

[0080] Para describir un segundo problema potencial, se observa que la fila 3 de la Tabla 10 de ISO/IEC 14496-15 establece:

'hvc2' o 'hev2'	Solo configuración HEVC	Una pista HEVC simple sin unidades NAL con nuh_layer_id mayor que 0; pueden estar presentes y usarse extractores para hacer referencia a unidades NAL; pueden estar presentes agregadores para contener y hacer referencia a unidades NAL.

[0081] Cuando el tipo de entrada de muestra es 'hvc2' o 'hev2', y la misma entrada contiene solo la configuración HEVC, puede ocurrir cualquiera de los siguientes escenarios: (1) La pista transporta un flujo de bits HEVC de una sola capa completa en el que todas las unidades NAL VCL tienen un nuh layer id igual a 0, y hay algunos extractores y/o agregadores presentes; o (2) La pista transporta un subconjunto de un flujo de bits HEVC de una sola capa de este tipo y el subconjunto contiene unidades NAL VCL que tienen solo un TemporalId mayor que 0 (independientemente de si hay extractores y/o agregadores).

[0082] Sin embargo, ninguno de los tipos de entrada de muestra 'hvc2' y 'hev2' se ha limitado para permitir un acceso aleatorio conveniente de manera similar al tipo de entrada de muestra 'hev1'. Esto significa que, para los dos escenarios anteriores, no hay forma de garantizar e indicar que el acceso aleatorio conveniente (sin necesidad de buscar y recuperar conjuntos de parámetros de muestras anteriores) esté habilitado.

[0083] Esta divulgación también describe técnicas que pueden usarse para resolver el segundo problema potencial. En particular, se puede especificar una restricción para permitir un acceso aleatorio conveniente, sin necesidad de buscar y recuperar conjuntos de parámetros de muestras anteriores para una pista que contenga solo un flujo de bits HEVC de capa única completa o un subconjunto del mismo, que el tipo de entrada de muestra 'hev2' se debe utilizar cuando la entrada de muestra contiene solo la configuración HEVC. Por lo tanto, cuando la entrada de muestra contiene solo la configuración HEVC, el dispositivo de preparación de contenido 20 puede especificar un tipo de entrada de muestra de 'hev2'. Del mismo modo, el dispositivo cliente 40 puede determinar que el acceso aleatorio conveniente está habilitado para una muestra que tiene el tipo de entrada de muestra 'hev2' y, además, que la muestra incluye datos de vídeo codificados de acuerdo solo con la configuración HEVC, en lugar de la configuración L-HEVC.

[0084] Para describir un tercer problema potencial, se observa que, cuando cada capa completa de un flujo de bits L-HEVC multicapa se almacena en una pista separada, de acuerdo con la especificación actual, es posible que la pista base use el tipo de entrada de muestra 'hevl' solo con la configuración HEVC, y otras pistas usan el tipo de entrada de muestra 'lhv1' solo con la configuración L-HEVC, y también es posible que la pista base use el tipo de entrada de muestra 'hvc1' solo con la configuración HEVC, y otras pistas usen el tipo de entrada de muestra 'lhe1' solo con la configuración L-HEVC.

[0085] En el primer escenario, se indica un acceso aleatorio conveniente para la pista base, pero no para otras pistas. En el segundo escenario, el acceso aleatorio conveniente no está indicado para la pista base, pero está indicado para otras pistas. Puede haber una excusa plausible para el primer escenario, como que la pista base es lo suficientemente importante como para justificar la sobrecarga de permitir un acceso aleatorio conveniente, mientras que la concesión de renunciar a un acceso aleatorio conveniente para a una menor sobrecarga para las pistas que llevan capas de mejora se considera una buena compensación. Sin embargo, el segundo escenario no tiene sentido ya que habilitar el acceso aleatorio conveniente para las pistas que llevan capas de mejora requeriría efectivamente habilitar el acceso aleatorio conveniente para la pista base, y si está habilitado para la pista base, no hay razón para no indicarlo usando el tipo de entrada de muestra correcto, es decir, 'hevl'.

[0086] De manera similar, el tercer problema potencial también puede aplicarse cuando un flujo de bits HEVC de una sola capa que contiene múltiples subcapas temporales es transportado por múltiples pistas, donde la pista que transporta la subcapa más baja (las unidades NAL VCL que tienen TemporalId igual a 0) usa un tipo de entrada de muestra que indica que el acceso aleatorio conveniente está habilitado, mientras que otra pista usa un tipo de entrada de muestra que no indica que el acceso aleatorio conveniente está habilitado, o viceversa.

[0087] Las técnicas de esta divulgación también pueden abordar el tercer problema potencial. En particular, se puede especificar una restricción que requiera que, para todas las pistas que transportan un flujo de bits de una o varias capas (codificado por HEVC, L-HEVC o cualquier otro códec), todas las pistas usan los tipos de entrada de muestra (junto con la presencia de configuraciones HEVC y/o L-HEVC) que indican que el acceso aleatorio conveniente está habilitado o todas las pistas usan los tipos de entrada de muestra (junto con la presencia de configuraciones HEVC y/o L-HEVC) que no indican que el acceso aleatorio conveniente está habilitado. Por lo tanto, el dispositivo de preparación de contenido 20 puede garantizar, para las pistas que transportan un flujo de bits de una o varias capas, que todas las pistas usan los tipos de entrada de muestra que indican que el acceso aleatorio conveniente está habilitado, o que todas las pistas usan tipos de entrada de muestra que indican que el acceso aleatorio conveniente no está habilitado. De esta manera, si el acceso aleatorio conveniente está habilitado para una de las pistas, el dispositivo cliente 40 puede determinar que el acceso aleatorio conveniente está habilitado para cada una de las pistas.

[0088] De forma alternativa, se puede especificar una restricción que requiera que, para todas las pistas que transportan un flujo de bits de una o varias capas (codificado por HEVC, L-HEVC o cualquier otro códec), cuando la pista que transporta la subcapa más baja (las unidades NAL VCL que tienen TemporalId igual a 0) de la capa base usa un tipo de entrada de muestra (junto con la presencia de configuraciones HEVC y/o L-HEVC) que indica que el acceso aleatorio conveniente está habilitado, todas las demás pistas deben usar tipos de entrada de muestra (junto con la presencia de configuraciones HEVC y/o L-HEVC) que indican que el acceso aleatorio conveniente está habilitado. Por lo tanto, el dispositivo de preparación de contenido 20 puede garantizar, para las pistas que transportan un flujo de bits de una o varias capas, que cuando una pista que transporta una subcapa temporal más baja de la capa base usa un tipo de entrada de muestra que indica que el acceso aleatorio conveniente está habilitado, que todas las pistas usen el tipo de entrada de muestra que indica que el acceso aleatorio conveniente está habilitado, o de forma alternativa, que si el acceso aleatorio conveniente no está habilitado para la subcapa temporal más baja de la capa base, que ninguna de las pistas use tipos de entrada de muestra que indican que el acceso aleatorio conveniente está habilitado. De esta manera, si el acceso aleatorio conveniente está habilitado para una pista que incluye la subcapa temporal más baja de la capa base, el dispositivo cliente 40 puede determinar que el acceso aleatorio conveniente está habilitado para cada una de las pistas.

[0089] El dispositivo cliente 40 puede usar las técnicas de esta divulgación para realizar un acceso aleatorio conveniente de varias maneras. Por ejemplo, el dispositivo cliente 40 puede realizar un acceso aleatorio conveniente solicitando datos del dispositivo servidor 60 usando un protocolo de red de transmisión en continuo, como DASH. En otros ejemplos, el dispositivo cliente 40 puede usar las técnicas de esta divulgación para realizar un acceso aleatorio conveniente para recuperar datos de medios de un medio de almacenamiento fijo legible por ordenador, como un disco versátil digital, disco Blu-ray, disco duro, memoria flash o similar. Por lo tanto, aunque la FIG. 1 ilustra un ejemplo que incluye transmisión en continuo basada en red, se debe entender que las técnicas de esta divulgación también se pueden aplicar en otros escenarios y contextos.

[0090] La FIG. 2 es un diagrama de bloques que ilustra con mayor detalle un conjunto de componentes de ejemplo de la unidad de recuperación 52 de la FIG. 1. En este ejemplo, la unidad de recuperación 52 incluye la unidad de middleware eMBMS 100, el cliente DASH 110 y la aplicación de medios 112.

[0091] En este ejemplo, la unidad de middleware eMBMS 100 incluye, además, la unidad de recepción eMBMS 106, la memoria caché 104 y la unidad servidor 102. En este ejemplo, la unidad de recepción eMBMS 106 está configurada para recibir datos a través de eMBMS, por ejemplo, de acuerdo con la entrega de archivos sobre transporte unidireccional (FLUTE), descrita en T. Paila y otros, "FLUTE-File Delivery over Unidirectional Transport", Grupo de trabajo de red, RFC 6726, noviembre de 2012, disponible en http://tools.ietf.org/html/rfc6726. Es decir, la unidad de recepción eMBMS 106 puede recibir archivos a través de radiodifusión desde, por ejemplo, el dispositivo servidor 60, que puede actuar como un BM-SC.

[0092] Dado que la unidad de middleware eMBMS 100 recibe los datos de los archivos, la unidad de middleware eMBMS puede almacenar los datos recibidos en la memoria caché 104. La memoria caché 104 puede comprender un medio de almacenamiento legible por ordenador, tal como memoria flash, un disco duro, RAM o cualquier otro medio de almacenamiento adecuado.

[0093] La unidad de servidor local 102 puede actuar como un servidor para el cliente DASH 110. Por ejemplo, la unidad de servidor local 102 puede proporcionar un archivo MPD u otro archivo de manifiesto al cliente DASH 110. La unidad de servidor local 102 anunciaría tiempos de disponibilidad para segmentos en el archivo MPD, así como hipervínculos desde los cuales los segmentos pueden recuperarse. Estos hipervínculos pueden incluir un prefijo de

dirección de localhost correspondiente al dispositivo cliente 40 (por ejemplo, 127.0.0.1 para IPv4). De esta manera, el cliente DASH 110 puede solicitar segmentos de la unidad de servidor local 102 usando peticiones GET o GET parcial HTTP. Por ejemplo, para un segmento disponible desde el enlace http://127.0.0.1/repl/seg3, el cliente DASH 110 puede construir una petición GET HTTP que incluya una petición para http://127.0.0. 1/rep1/seg3, y envíe la petición a la unidad de servidor local 102. La unidad de servidor local 102 puede recuperar los datos solicitados de la memoria caché 104 y proporcionar los datos al cliente DASH 110 en respuesta a tales peticiones.

[0094] La FIG. 3 es un diagrama conceptual que ilustra elementos del contenido multimedia 120 a modo de ejemplo. El contenido multimedia 120 puede corresponder al contenido multimedia 64 (FIG. 1), o a otro contenido multimedia almacenado en el medio de almacenamiento 62. En el ejemplo de la FIG. 3, el contenido multimedia 120 incluye una descripción de presentación de medios (MPD) 122 y una pluralidad de representaciones 124A-124N (representaciones 124). La representación 124A incluye datos de cabecera 126 y segmentos 128A-128N (segmentos 128) opcionales, mientras que la representación 124N incluye datos de cabecera 130 y segmentos 132A- 132N (segmentos 132) opcionales. La letra N se usa para designar, por razones de conveniencia, el último fragmento de película en cada una de las representaciones 124. En algunos ejemplos, puede haber diferentes números de fragmentos de películas entre las representaciones 124.

[0095] La MPD 122 puede comprender una estructura de datos separada de las representaciones 124. La MPD 122 puede corresponder al archivo de manifiesto 66 de la FIG. 1. Del mismo modo, las representaciones 124 pueden corresponder a las representaciones 68 de la FIG. 2. En general, la MPD 122 puede incluir datos que describan, en general, características de las representaciones 124, tales como las características de codificación y renderización, los conjuntos de adaptación, un perfil al que corresponda la MPD 122, la información del tipo de texto, la información del ángulo de la cámara, la información de calificación, la información del modo de avance y retroceso rápidos (por ejemplo, información indicativa de representaciones que incluyan subsecuencias temporales) y/o la información para recuperar períodos remotos (por ejemplo, para la inserción de anuncios dirigidos en el contenido de medios durante la reproducción).

[0096] Los datos de cabecera 126, cuando están presentes, pueden describir características de los segmentos 128, por ejemplo, localizaciones temporales de puntos de acceso aleatorio (RAP, también denominados puntos de acceso de flujo (SAP)), cuáles de los segmentos 128 incluyen puntos de acceso aleatorio, desplazamientos de bytes a puntos de acceso aleatorio dentro de los segmentos 128, localizadores de recursos uniformes (URL) de los segmentos 128 u otros aspectos de los segmentos 128. Los datos de cabecera 130, cuando están presentes, pueden describir características similares para los segmentos 132. Adicionalmente o de forma alternativa, dichas características pueden estar incluidas por completo dentro de la MPD 122.

[0097] Los segmentos 128, 132 incluyen una o más muestras de vídeo codificadas, cada una de las cuales puede incluir tramas o fragmentos de datos de vídeo. Cada una de las muestras de vídeo codificadas de los segmentos 128 puede tener características similares, por ejemplo, requisitos de altura, anchura y ancho de banda. Dichas características pueden describirse por datos de la MPD 122, aunque dichos datos no se ilustren en el ejemplo de la FIG. 3. La MPD 122 puede incluir características como se describen en la especificación 3GPP, con la adición de cualquiera o toda la información señalizada descrita en esta divulgación.

[0098] Cada uno de los segmentos 128, 132 puede estar asociado a un único localizador de recursos uniforme (URL). Por tanto, cada uno de los segmentos 128, 132 puede ser independientemente recuperable usando un protocolo de red de transmisión en continuo, tal como DASH. De esta manera, un dispositivo de destino, tal como el dispositivo cliente 40, puede usar una petición GET HTTP para recuperar los segmentos 128 o 132. En algunos ejemplos, el dispositivo cliente 40 puede usar peticiones GET parciales HTTP para recuperar intervalos de bytes específicos de los segmentos 128 o 132.

[0099] La FIG. 4 es un diagrama de bloques que ilustra elementos de un archivo de vídeo 150 de ejemplo, que puede corresponder a un segmento de una representación, tal como uno de los segmentos 114, 124 de la FIG. 3. Cada uno de los segmentos 128, 132 puede incluir datos que se ajustan sustancialmente a la disposición de datos ilustrada en el ejemplo de la FIG. 4. Se puede decir que el archivo de vídeo 150 encapsula un segmento. Como se ha descrito anteriormente, los archivos de vídeo, de acuerdo con el formato de archivo de medios de base ISO, y las ampliaciones del mismo, almacenan los datos en una serie de objetos, denominados "cajas". En el ejemplo de la FIG. 4, el archivo de vídeo 150 incluye la caja de tipo de archivo (FTYP) 152, la caja de película (MOOV) 154, las cajas de índices de segmento (sidx) 162, las cajas de fragmento de película (MOOF) 164 y la caja de acceso aleatorio de fragmento de película (MFRA) 166. Aunque la FIG. 4 representa un ejemplo de archivo de vídeo, se deberá entender que otros archivos de medios pueden incluir otros tipos de datos de medios (por ejemplo, datos de audio, datos de texto temporizado o similares) que están estructurados de forma similar a los datos del archivo de vídeo 150, de acuerdo con el formato de archivo de medios de base ISO y sus ampliaciones.

[0100] La caja de tipo de archivo (FTYP) 152 describe en general un tipo de archivo para el archivo de vídeo 150. La caja de tipo de archivo 152 puede incluir datos que identifican una especificación que describe un mejor uso para el archivo de vídeo 150. La caja de tipo de archivo 152 puede estar colocada de forma alternativa antes de la caja MOOV 154, las cajas de fragmento de película 164 y/o la caja MFRA 166.

[0101] En algunos ejemplos, un segmento, tal como el archivo de vídeo 150, puede incluir una caja de actualización de MPD (no mostrada) antes de la caja FTYP 152. La caja de actualización de MPD puede incluir información que indica que se va a actualizar una MPD correspondiente a una representación que incluye el archivo de vídeo 150, junto con información para actualizar la MPD. Por ejemplo, la caja de actualización de MPD puede proporcionar un URI o URL para un recurso que se va a usar para actualizar la MPD. Como otro ejemplo, la caja de actualización de MPD puede incluir datos para actualizar la MPD. En algunos ejemplos, la caja de actualización de MPD puede seguir inmediatamente a una caja de tipo de segmento (STYP) (no mostrada) del archivo de vídeo 150, donde la caja STYP puede definir un tipo de segmento para el archivo de vídeo 150. La FIG. 7, analizada con mayor detalle a continuación, proporciona información adicional con respecto a la caja de actualización de MPD.

[0102] La caja de MOOV 154, en el ejemplo de la FIG. 4, incluye la caja de cabecera de película (MVHD) 156, la caja de pista (TRAK) 158 y una o más cajas de ampliación de película (MVEX) 160. En general, la caja de MVHD 156 puede describir características generales del archivo de vídeo 150. Por ejemplo, la caja de MVHD 156 puede incluir datos que describen cuándo se creó inicialmente el archivo de vídeo 150, cuándo se modificó por última vez el archivo de vídeo 150, una escala de tiempo para el archivo de vídeo 150, una duración de reproducción para el archivo de vídeo 150 u otros datos que describen en general el archivo de vídeo 150.

[0103] La caja de TRAK 158 puede incluir datos para una pista del archivo de vídeo 150. La caja de TRAK 158 puede incluir una caja de cabecera de pista (TKHD) que describe las características de la pista correspondiente a la caja de TRAK 158. En algunos ejemplos, la caja TRAK 158 puede incluir imágenes de vídeo codificadas, mientras que, en otros ejemplos, las cajas de vídeo codificado de la pista pueden estar incluidas en fragmentos de película 164, a las cuales se puede hacer referencia mediante unos datos de la caja TRAK 158 y/o las cajas SIDX 162.

[0104] En algunos ejemplos, el archivo de vídeo 150 puede incluir más de una pista. Por consiguiente, la caja de MOOV 154 puede incluir un número de cajas de TRAK igual al número de pistas del archivo de vídeo 150. La caja de TRAK 158 puede describir las características de una pista correspondiente del archivo de vídeo 150. Por ejemplo, la caja de TRAK 158 puede describir información temporal y/o espacial para la pista correspondiente. Una caja de TRAK similar a la caja de TRAK 158 de la caja MOOV 154 puede describir características de una pista de conjunto de parámetros, cuando la unidad de encapsulación 30 (FIG. 3) incluye una pista de conjunto de parámetros en un archivo de vídeo, tal como el archivo de vídeo 150. La unidad de encapsulación 30 puede señalizar la presencia de mensajes de SEI a nivel de secuencia en la pista de conjunto de parámetros dentro del cuadro de TRAK que describe la pista de conjunto de parámetros.

[0105] Las cajas de MVEX 160 pueden describir características de correspondientes fragmentos de película 164, por ejemplo, para señalizar que el archivo de vídeo 150 incluye fragmentos de película 164, además de los datos de vídeo incluidos dentro de la caja de MOOV 154, si los hubiera. En el contexto de la transmisión continua de datos de vídeo, las imágenes de vídeo codificadas pueden estar incluidas en los fragmentos de película 164 en lugar de en la caja de MOOV 154. En consecuencia, todas las muestras de vídeo codificadas pueden estar incluidas en fragmentos de película 164, en lugar de en la caja de MOOV 154.

[0106] La caja de MOOV 154 puede incluir un número de cajas de MVEX 160 igual al número de fragmentos de película 164 del archivo de vídeo 150. Cada una de las cajas de MVEX 160 puede describir características de un fragmento correspondiente de los fragmentos de película 164. Por ejemplo, cada caja de MVEX puede incluir una caja de cabecera de ampliación de película (MEHD) que describe una duración temporal para la caja correspondiente de los fragmentos de película 164.

[0107] Como se indica anteriormente, la unidad de encapsulación 30 puede almacenar un conjunto de datos de secuencia en una muestra de vídeo que no incluye datos de vídeo codificados reales. Una muestra de vídeo puede corresponder en general a una unidad de acceso, que es una representación de una imagen codificada en una instancia de tiempo específica. En el contexto de la AVC, la imagen codificada incluye una o más unidades NAL VCL que contienen la información para construir todos los píxeles de la unidad de acceso y otras unidades NAL no VCL asociadas, tales como mensajes de SEI. Por consiguiente, la unidad de encapsulación 30 puede incluir un conjunto de datos de secuencia, que puede incluir mensajes de SEI a nivel de secuencia, en uno de los fragmentos de película 164. La unidad de encapsulación 30 puede señalizar, además, la presencia de un conjunto de datos de secuencia y/o de mensajes de SEI a nivel de secuencia con la presencia de estos en uno de los fragmentos de película 164 dentro de una de las cajas de MVEX 160 correspondiente al uno de los fragmentos de película 164.

[0108] Las cajas de SIDX 162 son elementos opcionales del archivo de vídeo 150. Es decir, los archivos de vídeo que se ajustan al formato de archivo 3GPP u otros formatos de archivo de este tipo, no incluyen necesariamente cajas de SIDX 162. De acuerdo con el ejemplo del formato de archivo 3GPP, se puede usar una caja de SIDX para identificar un subsegmento de un segmento (por ejemplo, un segmento contenido dentro del archivo de vídeo 150). El formato de archivo 3GPP define un subsegmento como "un conjunto autónomo de una o más cajas de fragmento de película consecutivas con un(as) caja(s) de datos de medios correspondiente(s), y una caja de datos de medios que contiene datos a los que se hace referencia mediante una caja de fragmento de película debe seguir a esa caja de fragmento de película y preceder a la siguiente caja de fragmento de película que contiene información sobre la misma pista". El

formato de archivo 3GPP también indica que una caja de SIDX "contiene una secuencia de referencias a subsegmentos del (sub)segmento documentado por la caja. Los subsegmentos a los que se hace referencia son contiguos en el tiempo de presentación. De forma similar, los bytes a los que una caja de índice de segmento hace referencia siempre son contiguos dentro del segmento. El tamaño al que se hace referencia da el recuento del número de bytes en el material al que se hace referencia".

[0109] Las cajas de SIDX 162 en general proporcionan información representativa de uno o más subsegmentos de un segmento incluido en el archivo de vídeo 150. Por ejemplo, dicha información puede incluir tiempos de reproducción en los que comienzan y/o terminan los subsegmentos, desplazamientos de bytes para los subsegmentos, si los subsegmentos incluyen (por ejemplo, comienzan con) un punto de acceso de flujo (SAP), un tipo para el SAP (por ejemplo, si el SAP es una imagen de actualización de descodificador instantánea (IDR), una imagen de acceso aleatorio limpio (CRA), una imagen de acceso de enlace roto (BLA) o similares), una posición del SAP (en términos de tiempo de reproducción y/o desplazamiento de bytes) en el subsegmento y similares.

[0110] Los fragmentos de película 164 pueden incluir una o más imágenes de vídeo codificadas. En algunos ejemplos, los fragmentos de película 164 pueden incluir uno o más grupos de imágenes (GOP), cada uno de los cuales puede incluir un número de imágenes de vídeo codificadas, por ejemplo, tramas o imágenes. Así mismo, como se describe anteriormente, los fragmentos de película 164 pueden incluir conjuntos de datos de secuencia en algunos ejemplos. Cada uno de los fragmentos de película 164 puede incluir una caja de cabecera de fragmento de película (MFHD, no mostrada en la FIG. 4). La caja MFHD puede describir características del fragmento de película correspondiente, tales como un número de secuencia para el fragmento de película. Los fragmentos de película 164 pueden estar incluidos por orden de número de secuencia en el archivo de vídeo 150.

[0111] La caja de MFRA 166 puede describir puntos de acceso aleatorio dentro de fragmentos de película 164 del archivo de vídeo 150. Esto puede ayudar a realizar modos de avance, tales como realizar búsquedas hasta ubicaciones temporales en particular (es decir, tiempos de reproducción) dentro de un segmento encapsulado por el archivo de vídeo 150. La caja de MFRA 166 en general es opcional y no necesita estar incluida en los archivos de vídeo, en algunos ejemplos. Del mismo modo, un dispositivo cliente, tal como el dispositivo cliente 40, no tiene necesariamente que hacer referencia a la caja de MFRA 166 para descodificar y visualizar correctamente los datos de vídeo del archivo de vídeo 150. La caja MFRA 166 puede incluir un número de cajas de acceso aleatorio de fragmento de pista (TFRA) (no mostradas) igual al número de pistas del archivo de vídeo 150 o, en algunos ejemplos, igual al número de pistas de medios (por ejemplo, pistas sin indicaciones) del archivo de vídeo 150.

[0112] En algunos ejemplos, los fragmentos de película 164 pueden incluir uno o más puntos de acceso de flujo (SAP), tales como imágenes IDR. Del mismo modo, la caja de MFRA 166 puede proporcionar indicaciones de ubicaciones dentro del archivo de vídeo 150 de los SAP. En consecuencia, se puede formar una subsecuencia temporal del archivo de vídeo 150 a partir de los SAP del archivo de vídeo 150. La subsecuencia temporal también puede incluir otras imágenes, tales como tramas P y/o tramas B que dependen de los SAP. Las tramas y/o fragmentos de la subsecuencia temporal pueden estar dispuestos dentro de los segmentos de modo que las tramas/fragmentos de la subsecuencia temporal que dependen de otras tramas/fragmentos de la subsecuencia pueden descodificarse apropiadamente. Por ejemplo, en la disposición jerárquica de los datos, los datos usados para la predicción de otros datos también pueden estar incluidos en la subsecuencia temporal.

[0113] El archivo de vídeo 150 también contiene la caja de descripción de muestra 168, en este ejemplo. En particular, la caja de descripción de muestra 168 se incluye dentro de la caja de TRAK 158, en este ejemplo. Una caja de descripción de muestra 168 de ejemplo se puede definir de la siguiente manera:

[0114] Entrada de muestra y tipos de cajas: ’hvc2’, ’hev2’, ’lhv1’, ’lhe1’, ’lhvC’

- Contenedor: Caja de descripción de muestra ('stsd')
- Obligatorio: Es obligatoria una entrada de muestra ’hvc1’, ’hev1’, ’hvc2’, ’hev2’, ’lhv1’, o ’lhe1’
- Cantidad: Pueden estar presentes una o más entradas de muestra

[0115] En esta definición de ejemplo para la caja de descripción de muestra 168, cuando el nombre de entrada de muestra es 'lhv1', el valor predeterminado y obligatorio de array_completeness es 1 para matrices de todos los tipos de conjuntos de parámetros, y 0 para todas las demás matrices. Cuando el nombre de entrada de muestra es 'lhe1', el valor predeterminado de array_completeness es 0 para todas las matrices.

[0116] En esta definición de ejemplo para la caja de descripción de muestra 168, cuando el nombre de entrada de muestra es 'hev2' y la entrada de muestra contiene solo la configuración HEVC, se aplican las mismas restricciones que se especifican en la cláusula 8.4.3 para el nombre de entrada de muestra 'hevl'.

[0117] En esta definición de ejemplo para la caja de descripción de muestra 168, cuando el nombre de entrada de muestra es 'lhe1', o cuando el nombre de entrada de muestra es 'hevl' o 'hev2' y la entrada de muestra contiene

configuraciones HEVC y L-HEVC, se aplica lo siguiente:

[0118] Las restricciones de ejemplo a continuación imponen restricciones en la colocación de conjuntos de parámetros fuera de banda (en entradas de muestra) y conjuntos de parámetros dentro de banda (en muestras), para permitir un acceso aleatorio conveniente desde unidades de acceso que contienen imágenes IRAP al menos en algunas capas. Con estas restricciones, un lector de archivos que se inicializa con las entradas de muestra y avanza desde una unidad de acceso en la que todas las imágenes son imágenes IRAP tendrá todos los conjuntos de parámetros que necesita.

[0119] En esta definición de ejemplo para la caja de descripción de muestra 168, para cualquier muestra particular en una pista particular, la muestra colocada temporalmente en otra pista se define como aquella con el mismo tiempo de descodificación que la de esta muestra particular.

[0120] En esta definición de ejemplo para la caja de descripción de muestra 168, para una imagen IRAP de una muestra, pista y capa dados, cada conjunto de parámetros necesario para descodificar la imagen IRAP se incluirá en uno de los siguientes:

a. la entrada de muestra que se aplica a la muestra dada en la pista dada

b. la entrada de muestra de la muestra inicial de una pista que lleva una capa de referencia de la capa dada, donde la muestra inicial es la muestra colocada temporalmente de la muestra dada, cuando la muestra colocada temporalmente contiene una imagen IRAP de la capa de referencia, o la muestra anterior que contiene una imagen IRAP de la capa de referencia

c. la muestra dada en sí, posiblemente mediante el uso de extractores d. cuando están presentes, cualquier muestra colocada temporalmente de las pistas que llevan capas de referencia de la capa dada, posiblemente mediante el uso de extractores

[0121] En esta definición de ejemplo para la caja de descripción de muestra 168, para una imagen no IRAP de una muestra, pista y capa dados, cada conjunto de parámetros necesario para descodificar esa imagen se incluirá en uno de los siguientes:

a. la entrada de muestra que se aplica a la muestra dada en la pista dada

b. la entrada de muestra de la muestra inicial de una pista que lleva una capa de referencia de la capa dada, donde la muestra inicial es la muestra colocada temporalmente de la muestra dada, cuando la muestra colocada temporalmente contiene una imagen IRAP de la capa de referencia, o la muestra anterior que contiene una imagen IRAP de la capa de referencia

c. cualquiera de las muestras en la pista dada desde la muestra anterior que contiene una imagen IRAP en la capa dada a la propia muestra dada, inclusive, posiblemente mediante el uso de extractores

d. d. cuando están presentes, cualquiera de las muestras en una pista que lleva una capa de referencia de la capa dada desde la muestra colocada temporalmente de la muestra anterior que contiene una imagen IRAP en la capa dada hasta la muestra colocada temporalmente de la muestra dada, inclusive, posiblemente usando extractores

[0122] En esta definición de ejemplo para la caja de descripción de muestra 168, para un flujo de bits HEVC o L-HEVC transportado en más de una pista, cuando el nombre de entrada de muestra de la pista base es 'hvc1' o 'hvc2', el nombre de entrada de muestra de otras pistas será 'hvc2' o 'lhv1', y cuando el nombre de entrada de muestra de la pista base sea 'hevl' o 'hev2', el nombre de entrada de muestra de otras pistas será 'hev2' o 'lhe1'. La pista base es la pista con la subcapa temporal más baja (cuyas unidades NAL VCL tienen TemporalId igual a 0) de la capa base presente de forma nativa.

[0123] La restricción anterior, en este ejemplo, requiere que se habilite un acceso aleatorio conveniente y se indique para todas las pistas que transportan un flujo de bits HEVC o L-HEVC o ninguna de las pistas.

[0124] Si las muestras de una pista contienen una capa base compatible con HEVC, en este ejemplo se utilizará una entrada de muestra 'hvc1', 'hev1', 'hvc2' o 'hev2'. Aquí, la entrada contendrá inicialmente una caja de configuración HEVC, posiblemente seguida de una caja de configuración L-HEVC como se define a continuación. La caja de configuración HEVC documenta el perfil, la capa, el nivel y posiblemente también los conjuntos de parámetros de la capa base compatible con HEVC, según lo definido por HEVCDecoderConfiguration-Record. La caja de configuración L-HEVC posiblemente documenta conjuntos de parámetros de las capas de mejora compatibles con L-HEVC según lo definido por LHEVCDecoderConfigurationRecord, almacenado en la caja de configuración L-HEVC.

[0125] Si las muestras de una pista no contienen una capa base HEVC, se utilizará el tipo de entrada de muestra

'lhv1' o 'lhe1' y la entrada de muestra contendrá una caja de configuración L-HEVC, como se define a continuación. Esto incluye un HEVCDecoderConfigurationRecord, como se define en, por ejemplo, ISO/IEC 14496-15.

[0126] La FIG. 5 es un diagrama de flujo que ilustra un procedimiento de formación y envío de datos de media de ejemplo de acuerdo con las técnicas de esta divulgación. El procedimiento de la FIG. 5 puede realizarse, por ejemplo, mediante el dispositivo de preparación de contenido 20 de la FIG. 1 y/o el dispositivo de servidor 60 de la FIG. 1, un dispositivo que está configurado para realizar la funcionalidad atribuida al dispositivo de preparación de contenido 20 y al dispositivo de servidor 60, o similar. Para fines de explicación, las técnicas de la FIG. 5 se analizan con respecto tanto al dispositivo de preparación de contenido 20 como al dispositivo de servidor 60.

[0127] Inicialmente, la unidad de encapsulación 30 recibe datos de medios codificados de, por ejemplo, el codificador de audio 26 y/o el codificador de vídeo 28 (200). En este ejemplo, la unidad de encapsulación 30 encapsula datos de vídeo en varias muestras, algunas de las cuales representan muestras de acceso aleatorio. Los datos de vídeo pueden codificarse como HEVC o L-HEVC. Para permitir un acceso aleatorio conveniente, la unidad de encapsulación 30 asigna valores de tipo de entrada de muestra de 'hevl' o 'hev2' a muestras que encapsulan datos de medios de punto de acceso aleatorio (202), y también proporciona conjuntos de parámetros con la entrada de muestra o muestras de acceso aleatorio (204). Los puntos de acceso aleatorio pueden corresponder a tramas de actualización instantánea del descodificador (IDR), que pueden corresponder a tramas codificadas que usan intrapredicción (tramas I). Los conjuntos de parámetros pueden incluir cualquiera o todos los conjuntos de parámetros de vídeo (VPS), conjuntos de parámetros de secuencia (SPS) y/o conjuntos de parámetros de imagen (SPS). En general, la unidad de encapsulación 30 puede garantizar que todos los conjuntos de parámetros necesarios se proporcionen con las muestras de acceso aleatorio y/o con las entradas de muestra correspondientes a las muestras de acceso aleatorio, para permitir un acceso aleatorio conveniente.

[0128] La unidad de encapsulación 30 puede entonces formar uno o más archivos, por ejemplo, de acuerdo con ISO/IEC 14496-15, incluidas las muestras y los conjuntos de parámetros (206). En particular, la unidad de encapsulación 30 puede formar los archivos de modo que los conjuntos de parámetros se incluyan con los tipos de entrada de muestra 'hevl' o 'hev2' y/o las muestras correspondientes a estos tipos de entrada de muestra, para permitir un acceso aleatorio conveniente. El dispositivo de preparación de contenido 20 puede entonces generar datos de tipo de entrada de muestra para una muestra de acceso aleatorio (208). En algunos ejemplos, esta salida puede ser a un medio fijo, como un disco digital versátil (DVD) o un disco Blu-ray, junto con el resto del archivo o archivos. En otros ejemplos, esta salida puede enviarse al dispositivo servidor 60, que luego puede enviar los datos del tipo de entrada de muestra a un dispositivo cliente, tal como el dispositivo cliente 40 de la FIG. 1.

[0129] En el ejemplo de la FIG. 5, el dispositivo servidor 60 envía los datos del tipo de entrada de muestra al dispositivo cliente 40, lo que hace que el dispositivo cliente 40 realice un acceso aleatorio conveniente. Por lo tanto, la unidad de procesamiento de peticiones 70 del dispositivo servidor 60 de la FIG. 1 recibe una petición de una muestra de acceso aleatorio, para la cual un tipo de entrada de muestra es 'hevl' o 'hev2', del dispositivo cliente 40 (210). En respuesta, la unidad de procesamiento de peticiones 70 envía la muestra de acceso aleatorio solicitada y los conjuntos de parámetros (VPS, SPS y/o PPS) al dispositivo cliente 40 (212). En particular, estos datos pueden organizarse de modo que los conjuntos de parámetros se coloquen con los datos de entrada de muestra o con la muestra correspondiente a los datos de entrada de muestra, de modo que el dispositivo cliente 40 no necesite buscar y recuperar conjuntos de parámetros de muestras anteriores a la muestra de acceso aleatorio. De esta manera, las técnicas de la FIG. 5 puede permitir un acceso aleatorio conveniente para el dispositivo cliente 40, reduciendo así los requisitos de procesamiento para el dispositivo servidor 60 y el dispositivo cliente 40, reduciendo el consumo de ancho de banda para este intercambio de datos en comparación con si el acceso aleatorio conveniente no estuviera habilitado y mejorando la latencia entre un momento en el que el dispositivo cliente 40 recibe la entrada que solicita datos de medios de un usuario y el momento en que el dispositivo cliente 40 puede presentar los datos de medios solicitados al usuario.

[0130] La FIG. 6 es un diagrama de flujo que ilustra un procedimiento de realización de acceso aleatorio de ejemplo de acuerdo con las técnicas de esta divulgación. Para propósitos de ejemplo, el procedimiento según la FIG. 6 se explica con respecto al dispositivo cliente 40 de la FIG. 1.

[0131] En este ejemplo, la unidad de recuperación 52 del dispositivo cliente 40 solicita inicialmente datos de tipo de entrada de muestra para una muestra de punto de acceso aleatorio (220). Por ejemplo, la unidad de recuperación 52 puede enviar una petición HTTP para los datos del tipo de entrada de muestra de acuerdo con DASH. Después de recibir los datos del tipo de entrada de muestra, la unidad de procesamiento de formato de archivo 50 del dispositivo cliente 40 determina que los datos del tipo de entrada de muestra indican un tipo de entrada de muestra de 'hevl' o 'hev2' (222) para una muestra asociada que incluye datos de vídeo codificados de acuerdo con, por ejemplo, uno de HEVC o L-HEVC. En consecuencia, el dispositivo cliente 40 puede determinar que la muestra se puede usar para un acceso aleatorio conveniente. Es decir, el dispositivo cliente 40 puede recuperar la muestra y los conjuntos de parámetros incluidos con los datos de medios de la muestra o con la entrada de muestra para la muestra, sin buscar ni recuperar conjuntos de parámetros de muestras anteriores en orden de descodificación. Debe entenderse que, en este ejemplo, el flujo de bits de vídeo que incluye la muestra también incluye una o más muestras anteriores a la muestra en orden de descodificación.

[0132] En respuesta, el dispositivo cliente 40 recupera los datos de medios de muestra y los conjuntos de parámetros correspondientes (226). El dispositivo cliente 40 no necesita buscar y recuperar conjuntos de parámetros de muestras anteriores, sino que los conjuntos de parámetros recuperados incluyen todos los datos de conjunto de parámetros necesarios para descodificar los datos de medios de la muestra recuperada, debido al acceso aleatorio conveniente. Por lo tanto, la unidad de procesamiento de formato de archivo 50 desencapsula datos de vídeo de un archivo que incluye la muestra recuperada, y proporciona los datos de medios desencapsulados al descodificador de vídeo 48. El descodificador de vídeo 48 descodifica los datos de vídeo usando los conjuntos de parámetros, y proporciona los datos de vídeo descodificados a la salida de vídeo 44, que presenta los datos de medios (228). La unidad de recuperación 52 puede solicitar, además, una muestra posterior en el orden de descodificación (230), y el descodificador de vídeo 48 puede descodificar los datos de vídeo de la muestra posterior y la salida de vídeo 44 puede presentar los datos de vídeo descodificados. Este proceso puede continuar hasta el final de la presentación de medios correspondiente, o hasta que un usuario, por ejemplo, solicite un nuevo conjunto de datos de medios.

[0133] De esta manera, el procedimiento de la FIG. 6 representa un ejemplo de un procedimiento de procesamiento de datos de vídeo, que incluye recibir datos que describen un tipo de entrada de muestra para una muestra de un flujo de bits de vídeo, siendo el tipo de entrada de muestra uno de 'hevl' o 'hev2', en el que la muestra comprende datos de vídeo codificados de acuerdo con una codificación de vídeo de alta eficiencia (HEVC) o HEVC en capas (L-HEVC), y en el que una o más muestras adicionales, incluidos los datos de vídeo, preceden a la muestra en el flujo de bits de vídeo en orden de descodificación, y en respuesta a que el tipo de entrada de muestra sea 'hevl' o 'hev2' y comprendiendo la muestra los datos de vídeo codificados de acuerdo con uno de HEVC o L-HEVC, recuperando la muestra para realizar un acceso aleatorio usando la muestra, sin recuperar los datos de vídeo de uno o más muestras que preceden a la muestra y sin recuperar conjuntos de parámetros de ninguna muestra previa del flujo de bits de vídeo en orden de descodificación.

[0134] La FIG. 7 es un diagrama de flujo que ilustra una técnica de ejemplo para generar un archivo que incluye datos de vídeo. El procedimiento de la FIG. 7 se describe como realizado por la unidad de encapsulación 30 de la FIG. 1. Sin embargo, debe entenderse que, en general, el procedimiento de la FIG. 7 también puede ser realizado por otros dispositivos, por ejemplo, por componentes del dispositivo servidor 60. Además, las técnicas de la FIG. 7 puede realizarse junto con las técnicas de generación de archivos de la FIG. 5 como se analizó anteriormente.

[0135] Inicialmente, la unidad de encapsulación 30 recibe datos de medios codificados (250). La unidad de encapsulación 30 puede recibir los datos de medios codificados de, por ejemplo, el codificador de vídeo 28, que puede codificar los datos de medios de acuerdo con HEVC, L-HEVC u otro estándar de codificación de vídeo similar. En particular, los datos de medios codificados pueden incluir una pluralidad de capas, por ejemplo, para MV-HEVC, 3D-HEVC, SHVC o similares (como extensiones en capas para otros estándares de codificación de vídeo).

[0136] La unidad de encapsulación 30 puede formar, en general, un archivo que incluye los datos de medios codificados en forma de una pluralidad de pistas. La unidad de encapsulación 30 puede formar pistas del archivo de manera que cada pista que incluya datos de vídeo incluya una o más capas de datos de vídeo de múltiples capas (252). Múltiples capas dentro de una pista pueden corresponder, por ejemplo, a varias capas de adaptabilidad a escala temporal. Las pistas que incluyen distintas capas pueden corresponder, por ejemplo, a distintas vistas para MV-HEVC o 3DHEVC, o diferentes capas de adaptabilidad a escala para SHVC (por ejemplo, adaptabilidad a escala de resolución espacial, adaptabilidad a escala de profundidad de bits, adaptabilidad a escala de la relación señal/ruido máxima (PSNR) o similares).

[0137] En el ejemplo de la FIG. 7, la unidad de encapsulación 30 determina a continuación si se debe habilitar el acceso aleatorio conveniente (254). Por ejemplo, la unidad de encapsulación 30 puede recibir datos de configuración de un usuario, como un administrador. De acuerdo con la restricción propuesta en esta divulgación, cuando se habilita el acceso aleatorio conveniente para una pista más baja de una pluralidad de pistas, la pista más baja incluye una capa base de datos de vídeo que lleva una subcapa más baja de los datos de vídeo, la unidad de encapsulación 30 permite un acceso aleatorio conveniente para cada pista de la pluralidad de pistas que incluye datos de vídeo (256). De forma alternativa, si el acceso aleatorio conveniente está deshabilitado para la pista más baja, la unidad de encapsulación 30 deshabilita el acceso aleatorio conveniente para las otras pistas de la pluralidad de pistas (258). En otros ejemplos, la unidad de encapsulación 30 puede habilitar o deshabilitar el acceso aleatorio conveniente de acuerdo con la restricción alternativa analizada anteriormente (es decir, para todas las pistas que transportan un flujo de bits de una o varias capas (codificado por HEVC, L-HEVC o cualquier otro códec), cuando la pista que lleva la subcapa más baja (cuyas unidades NAL VCL tienen TemporalId igual a 0) de la capa base usa un tipo de entrada de muestra (junto con la presencia de configuraciones HEVC y/o L-HEVC) que indica que el acceso aleatorio conveniente está habilitado, todas las demás pistas deben usar tipos de entrada de muestra (junto con la presencia de configuraciones HEVC y/o L-HEVC) que indican que el acceso aleatorio conveniente está habilitado).

[0138] La unidad de encapsulación 30 luego asigna valores de tipo de entrada de muestra que indican si el acceso aleatorio conveniente está habilitado para entradas de muestra para muestras de las pistas que incluyen datos de vídeo (260). Por ejemplo, 'hevl' y 'hev2' como valores de tipo de entrada de muestra pueden indicar que el acceso aleatorio conveniente está habilitado para las pistas que incluyen los datos de vídeo. La unidad de encapsulación 30

también forma un archivo que incluye las pistas y los valores del tipo de entrada de muestra como se determinó anteriormente (262).

[0139] De esta manera, el procedimiento de la FIG. 7 representa un ejemplo de un procedimiento para generar un archivo que incluye datos de vídeo que incluye, en respuesta a la determinación de que una pista más baja de una pluralidad de pistas, incluyendo la pista más baja una capa base de datos de vídeo que lleva una subcapa más baja de los datos de vídeo, debe incluir valores de tipo de entrada de muestra para muestras que indican que el acceso aleatorio conveniente está habilitado, establecer valores de tipo de entrada de muestra para muestras de cada una de las otras pistas de la pluralidad de pistas que incluyen datos de vídeo para indicar que el acceso aleatorio conveniente está habilitado, y generar un archivo que incluye la pluralidad de pistas, de modo que los valores de tipo de entrada de muestra para las pistas de la pluralidad de pistas indican que el acceso aleatorio conveniente está habilitado.

[0140] En uno o más ejemplos, las funciones descritas se pueden implementar en hardware, software, firmware o en cualquier combinación de los mismos. Si se implementan en software, las funciones se pueden almacenar en, o transmitir por, un medio legible por ordenador como una o más instrucciones o código, y ejecutar mediante una unidad de procesamiento basada en hardware. Los medios legibles por ordenador pueden incluir medios de almacenamiento legibles por ordenador, que correspondan a un medio tangible tal como medios de almacenamiento de datos, o medios de comunicación, incluyendo cualquier medio que facilite la transferencia de un programa informático de un lugar a otro, por ejemplo, de acuerdo con un protocolo de comunicación. De esta manera, los medios legibles por ordenador pueden corresponder, en general, a (1) medios de almacenamiento tangibles legibles por ordenador que sean no transitorios o a (2) un medio de comunicación tal como una señal o una onda portadora. Los medios de almacenamiento de datos pueden ser medios disponibles cualesquiera a los que se puede acceder desde uno o más ordenadores o uno o más procesadores para recuperar instrucciones, código y/o estructuras de datos para la implementación de las técnicas descritas en esta divulgación. Un producto de programa informático puede incluir un medio legible por ordenador.

[0141] A modo de ejemplo, y no de limitación, dichos medios de almacenamiento legibles por ordenador pueden comprender RAM, ROM, EEPROM, CD-ROM u otro almacenamiento en disco óptico, almacenamiento en disco magnético u otros dispositivos de almacenamiento magnético, memoria flash o cualquier otro medio que se pueda usar para almacenar el código de programa deseado en forma de instrucciones o estructuras de datos y al que se pueda acceder mediante un ordenador. Asimismo, cualquier conexión recibe apropiadamente la denominación de medio legible por ordenador. Por ejemplo, si las instrucciones se transmiten desde una página web, un servidor u otra fuente remota usando un cable coaxial, un cable de fibra óptica, un par trenzado, una línea de abonado digital (DSL) o tecnologías inalámbricas tales como infrarrojos, radio y microondas, entonces el cable coaxial, el cable de fibra óptica, el par trenzado, la DSL o las tecnologías inalámbricas tales como infrarrojos, radio y microondas están incluidos en la definición de medio. Sin embargo, se deberá entender que los medios de almacenamiento legibles por ordenador y los medios de almacenamiento de datos no incluyen conexiones, ondas portadoras, señales u otros medios transitorios, sino que, en cambio, están dirigidos a medios de almacenamiento no transitorios tangibles. Los discos, como se usan en el presente documento, incluyen el disco compacto (CD), el disco láser, el disco óptico, el disco versátil digital (DVD), el disco flexible y el disco Blu-ray, donde los discos flexibles reproducen normalmente datos magnéticamente, mientras que los demás discos reproducen los datos ópticamente con láseres. Las combinaciones de los anteriores también se deben incluir dentro del alcance de los medios legibles por ordenador.

[0142] Uno o más procesadores, tales como uno o más procesadores de señales digitales (DSP), microprocesadores de propósito general, circuitos integrados específicos de la aplicación (ASIC), matrices lógicas programables in situ (FPGA) u otros circuitos lógicos integrados o discretos equivalentes pueden ejecutar instrucciones. Por consiguiente, el término "procesador", como se usa en el presente documento, se puede referir a cualquiera de las estructuras anteriores o a cualquier otra estructura adecuada para la implementación de las técnicas descritas en el presente documento. Además, en algunos aspectos, la funcionalidad descrita en el presente documento se puede proporcionar dentro de módulos de hardware y/o de software dedicados configurados para codificar y descodificar, o incorporar en un códec combinado. Además, las técnicas se podrían implementar por completo en uno o más circuitos o elementos lógicos.

[0143] Las técnicas de la presente divulgación se pueden implementar en una amplia variedad de dispositivos o aparatos, incluyendo un teléfono inalámbrico, un circuito integrado (CI) o un conjunto de CI (por ejemplo, un conjunto de chips). En la presente divulgación se describen diversos componentes, módulos o unidades para destacar aspectos funcionales de dispositivos configurados para realizar las técnicas divulgadas, pero no se requiere necesariamente su realización mediante diferentes unidades de hardware. En su lugar, como se describe anteriormente, diversas unidades se pueden combinar en una unidad de hardware de códec o proporcionarse mediante un grupo de unidades de hardware interoperativas, que incluya uno o más procesadores como se describe anteriormente, junto con software y/o firmware adecuados.

[0144] Se han descrito diversos ejemplos. Estos y otros ejemplos están dentro del alcance de las siguientes reivindicaciones.

REIVINDICACIONES

1. Un procedimiento de recuperación de datos de vídeo, comprendiendo el procedimiento:

recibir datos que describen un tipo de entrada de muestra para una muestra de un flujo de bits de vídeo, siendo el tipo de entrada de muestra uno de 'hev1' o 'hev2', en el que la muestra comprende datos de vídeo codificados de acuerdo con una de codificación de vídeo de alta eficacia (HEVC) o HEVC en capas (L-HEVC), en el que una o más de otras muestras que incluyen datos de vídeo preceden a la muestra en el flujo de bits de vídeo en orden de descodificación, y

en respuesta a que el tipo de entrada de muestra es 'hev1' o 'hev2' y que la muestra comprende los datos de vídeo codificados de acuerdo con uno de HEVC o L-HEVC, recuperar la muestra para realizar un acceso aleatorio usando la muestra, sin recuperar los datos de vídeo de cualquiera de las una o más de otras muestras que preceden a la muestra, y sin recuperar conjuntos de parámetros de ninguna muestra anterior del flujo de bits de vídeo en orden de descodificación,

caracterizado por que los datos de vídeo de la muestra se incluyen en una pista de una pluralidad de pistas de un archivo, incluyendo la pluralidad de pistas un subconjunto de la pluralidad de pistas que incluye datos de vídeo de una o varias capas, comprendiendo además el procedimiento, en respuesta a determinar que el acceso aleatorio conveniente está habilitado para uno del subconjunto de la pluralidad de pistas que tiene una subcapa temporal más baja, determinar que el acceso aleatorio conveniente está habilitado para cada una de la pluralidad de pistas.

2. El procedimiento de la reivindicación 1, que comprende además recibir conjuntos de parámetros que describen la muestra en al menos una de una entrada de muestra correspondiente a la muestra o en la muestra.

3. El procedimiento de la reivindicación 1, en el que la muestra comprende los datos de vídeo de acuerdo con HEVC, y en el que el tipo de entrada de muestra es 'hev2'.

4. El procedimiento de la reivindicación 1, en el que los datos de vídeo de la muestra se incluyen en una pista de un archivo.

5. El procedimiento de la reivindicación 4, en el que el archivo incluye una pluralidad de pistas, cada una de las cuales corresponde a una capa de L-HEVC.

6. El procedimiento de la reivindicación 1, en el que los datos de vídeo de la muestra se incluyen en una pista de una pluralidad de pistas de un archivo, incluyendo la pluralidad de pistas un subconjunto de la pluralidad de pistas que incluye datos de vídeo de una o varias capas, comprendiendo además el procedimiento determinar que cada subconjunto de la pluralidad de pistas incluye muestras que tienen tipos de entrada de muestra que indican que el acceso aleatorio conveniente está habilitado.

7. Un dispositivo para recuperar datos de vídeo, comprendiendo el dispositivo:

medios para recibir datos que describen un tipo de entrada de muestra para una muestra de un flujo de bits de vídeo, siendo el tipo de entrada de muestra uno de 'hev1' o 'hev2', en el que la muestra comprende datos de vídeo codificados de acuerdo con una de codificación de vídeo de alta eficacia (HEVC) o HEVC en capas (L-HEVC), en el que una o más de otras muestras que incluyen datos de vídeo preceden a la muestra en el flujo de bits de vídeo en orden de descodificación,

medios para recuperar la muestra para realizar un acceso aleatorio usando la muestra en respuesta a que el tipo de entrada de muestra es 'hev1' o 'hev2' y que la muestra comprende los datos de vídeo codificados de acuerdo con uno de HEVC o L-HEVC, sin recuperar los datos de vídeo de cualquiera de las una o más de otras muestras que preceden a la muestra, y sin recuperar conjuntos de parámetros de ninguna muestra anterior del flujo de bits de vídeo en orden de descodificación,

caracterizado por que los datos de vídeo de la muestra se incluyen en una pista de una pluralidad de pistas de un archivo, incluyendo la pluralidad de pistas un subconjunto de la pluralidad de pistas que incluye datos de vídeo de una o varias capas; medios para determinar que el acceso aleatorio conveniente está habilitado para cada una de la pluralidad de pistas en respuesta a determinar que el acceso aleatorio conveniente está habilitado para uno del subconjunto de la pluralidad de pistas que tiene una subcapa temporal más baja.

8. El dispositivo de la reivindicación 7, que comprende además medios para recibir conjuntos de parámetros que describen la muestra en al menos una de una entrada de muestra correspondiente a la muestra o en la muestra.

9. El dispositivo de la reivindicación 7, en el que la muestra comprende los datos de vídeo de acuerdo con HEVC, y en el que el tipo de entrada de muestra es 'hev2'.

10. El dispositivo de la reivindicación 7, en el que los datos de vídeo de la muestra se incluyen en una pista de un archivo.

11. El dispositivo de la reivindicación 10, en el que el archivo incluye una pluralidad de pistas, cada una de las cuales corresponde a una capa de L-HEVC.

12. El dispositivo de la reivindicación 7, en el que los datos de vídeo de la muestra se incluyen en una pista de una pluralidad de pistas de un archivo, incluyendo la pluralidad de pistas un subconjunto de la pluralidad de pistas que incluye datos de vídeo de una o varias capas, comprendiendo además medios para determinar que cada subconjunto de la pluralidad de pistas incluye muestras que tienen tipos de entrada de muestra que indican que el acceso aleatorio conveniente está habilitado.

13. El dispositivo de la reivindicación 7, en el que el dispositivo es un dispositivo de comunicación inalámbrica, que comprende, además, un receptor configurado para recibir el archivo.

14. El dispositivo de la reivindicación 13, en el que el dispositivo de comunicación inalámbrica es un teléfono celular y el archivo se recibe por el receptor y se modula de acuerdo con un estándar de comunicación celular.

15. Un medio de almacenamiento legible por ordenador que tiene instrucciones almacenadas en el mismo que, cuando se ejecutan, hacen que un procesador realice el procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 6.

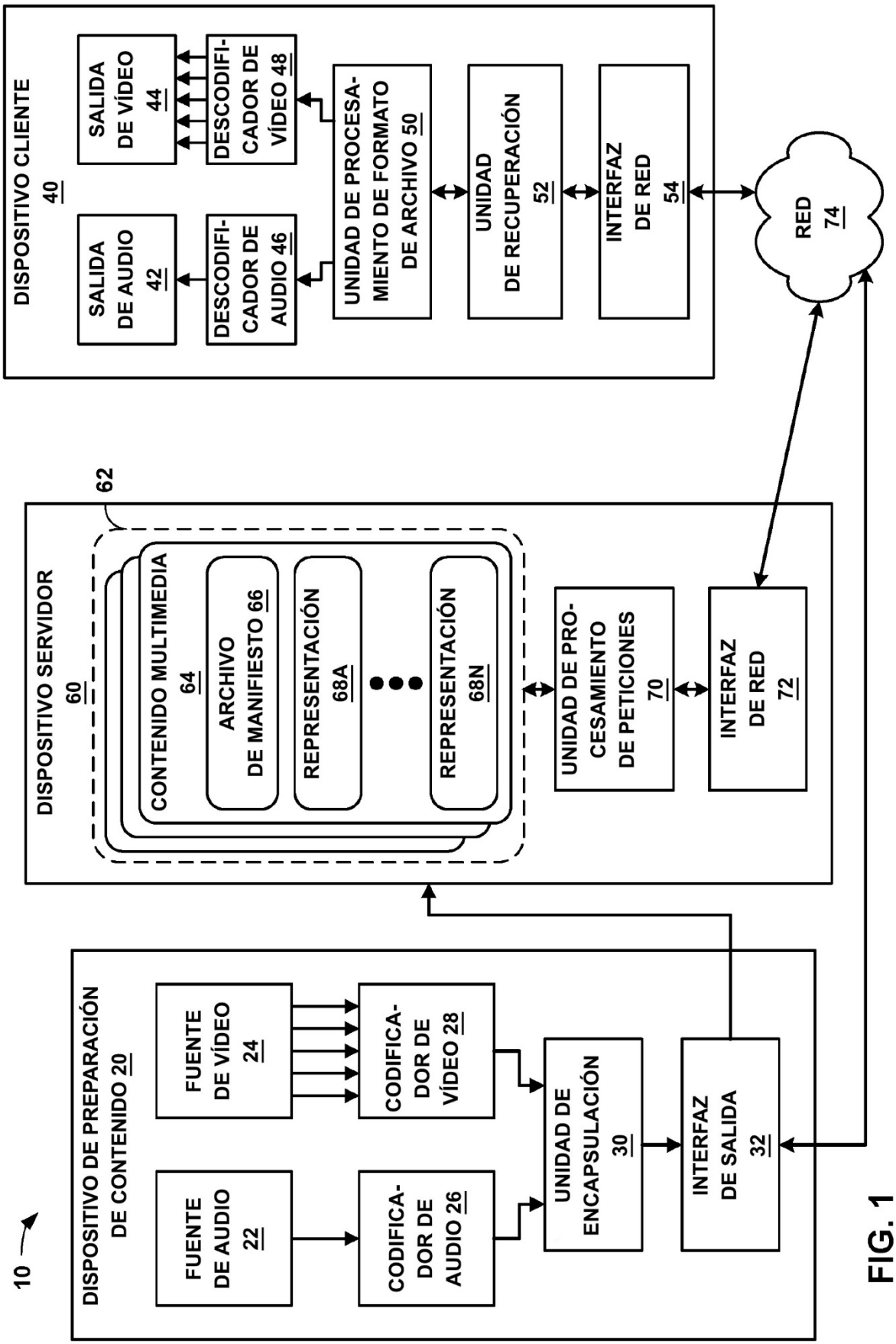
10
DISPOSITIVO DE PREPARACIÓN DE CONTENIDO 20
FUENTE DE AUDIO 22
FUENTE DE VÍDEO 24
CODIFICADOR DE AUDIO 26
CODIFICADOR DE VÍDEO 28
UNIDAD DE ENCAPSULACIÓN 30
INTERFAZ DE SALIDA 32
DISPOSITIVO SERVIDOR 60
62
CONTENIDO MULTIMEDIA 64
ARCHIVO DE MANIFIESTO 66
REPRESENTACIÓN 68A
REPRESENTACIÓN 68N
UNIDAD DE PROCESAMIENTO DE PETICIONES 70
INTERFAZ DE RED 72
RED 74
DISPOSITIVO CLIENTE 40
SALIDA DE AUDIO 42
SALIDA DE VÍDEO 44
DESCODIFICADOR DE AUDIO 46
DESCODIFICADOR DE VÍDEO 48
UNIDAD DE PROCESAMIENTO DE FORMATO DE ARCHIVO 50
UNIDAD DE RECUPERACIÓN 52
INTERFAZ DE RED 54

FIG. 1

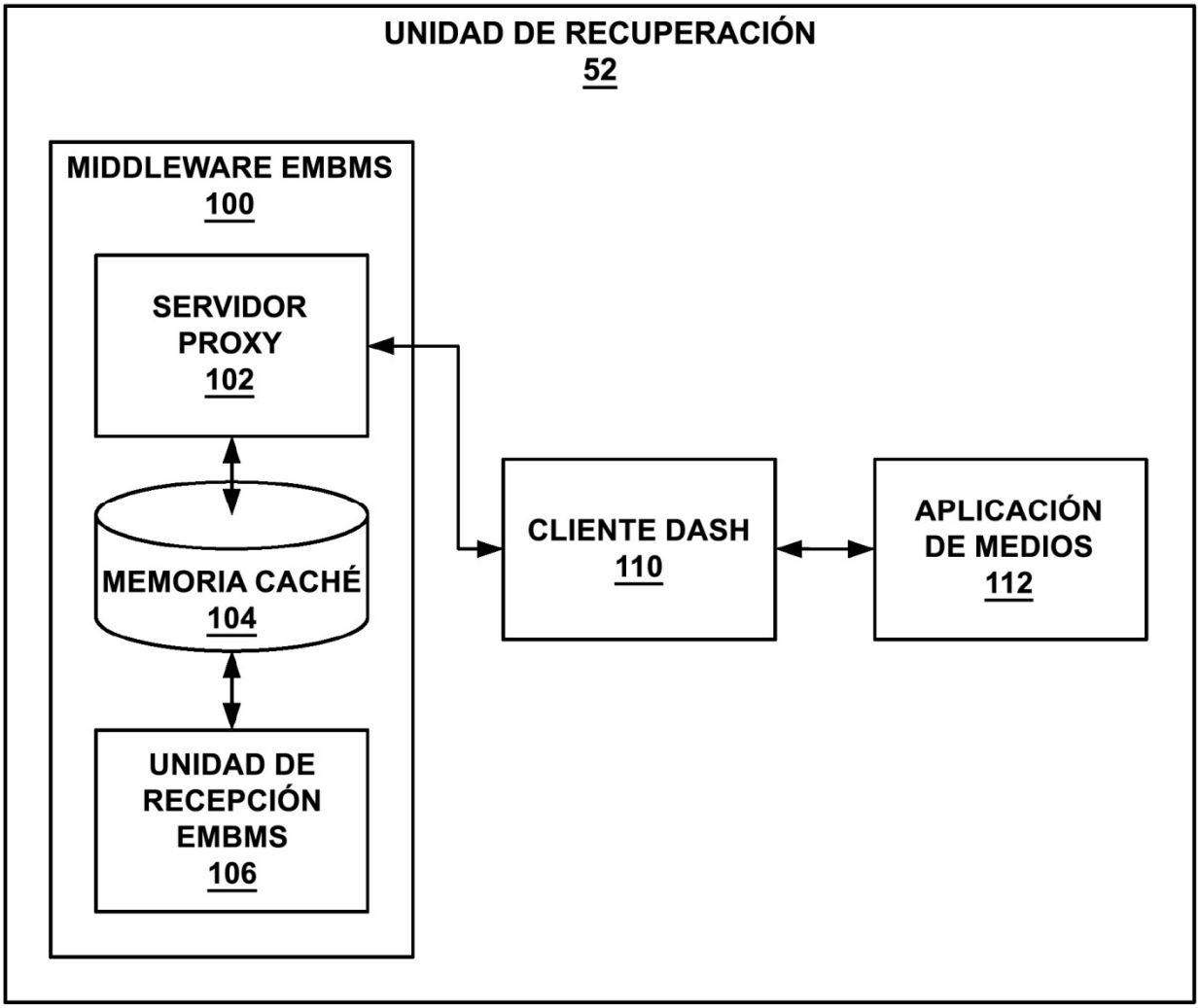
UNIDAD DE RECUPERACIÓN
52
MIDDLEWARE EMBMS
100
SERVIDOR PROXY
102
MEMORIA CACHÉ
104
UNIDAD DE RECEPCIÓN EMBMS
106
CLIENTE DASH
110
APLICACIÓN DE MEDIOS
112

FIG. 2

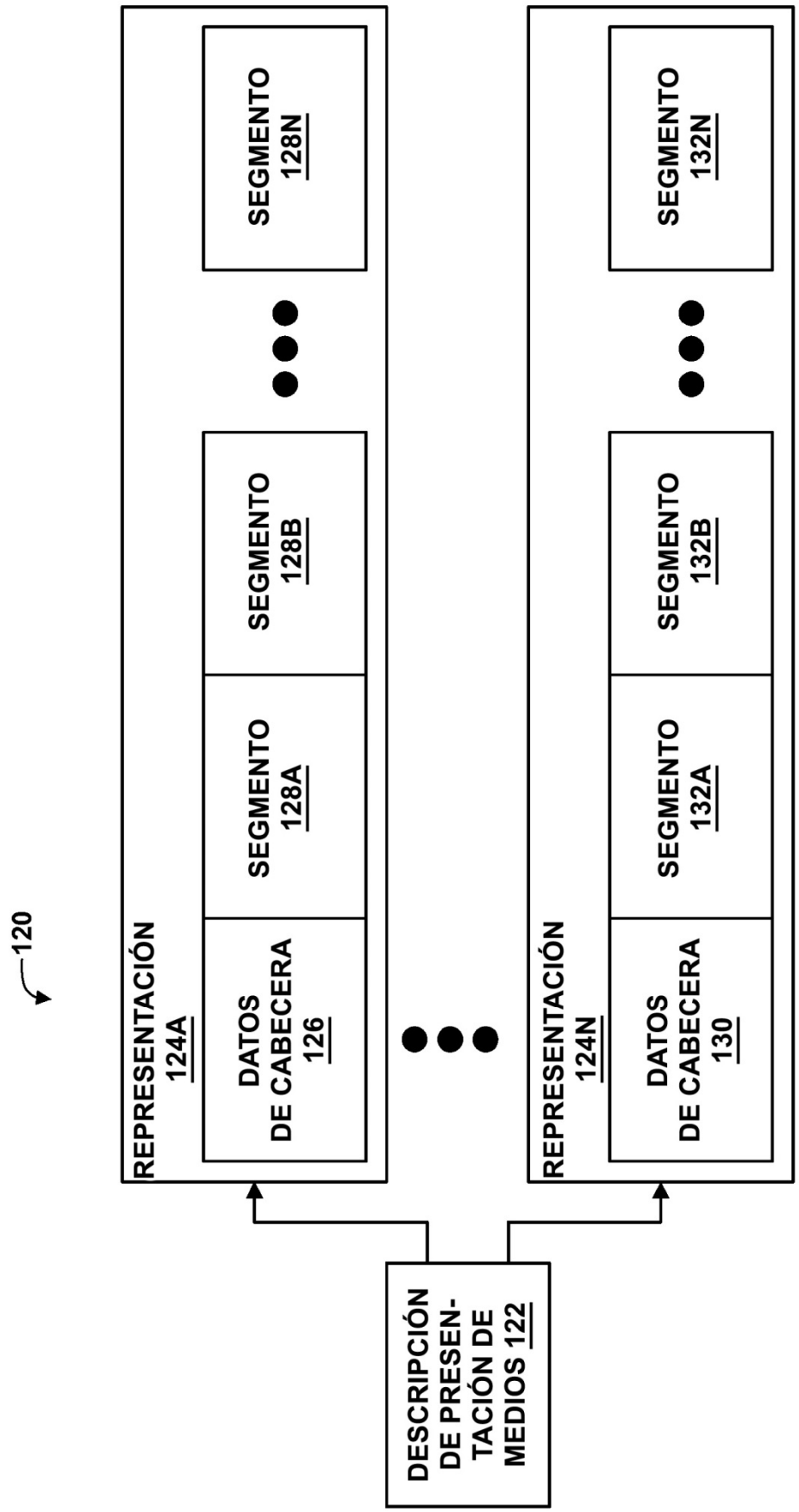
120
REPRESENTACIÓN 124A
DATOS DE CABECERA 126
SEGMENTO 128A
SEGMENTO 128B
SEGMENTO 128N
REPRESENTACIÓN 124N
DATOS DE CABECERA 130
SEGMENTO 132A
SEGMENTO 132B
SEGMENTO 132N
DESCRIPCIÓN DE PRESEN-TACIÓN DE MEDIOS 122

FIG. 3

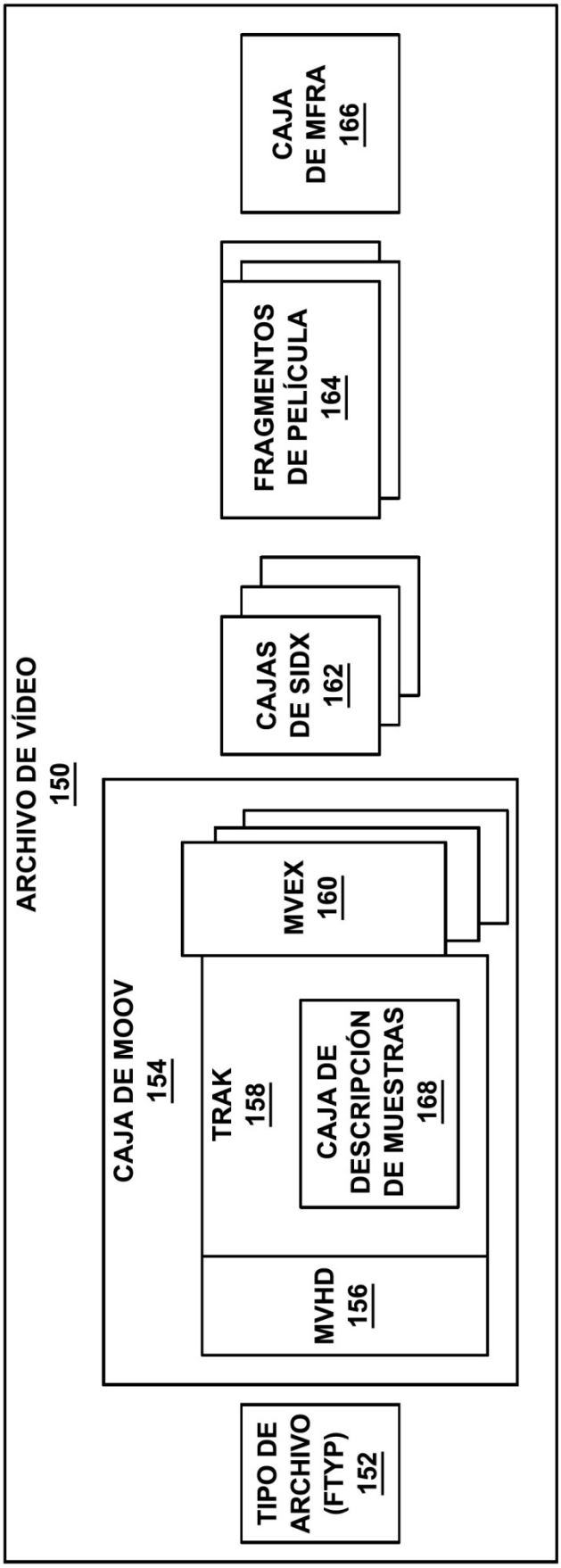
ARCHIVO DE VÍDEO
150
TIPO DE ARCHIVO (FTYP)
152
CAJA DE MOOV
154
MVHD
156
TRAK
158
CAJA DE DESCRIPCIÓN DE MUESTRAS
168
MVEX
160
CAJAS DE SIDX
162
FRAGMENTOS DE PELÍCULA
164
CAJA DE MFRA
166

FIG. 4

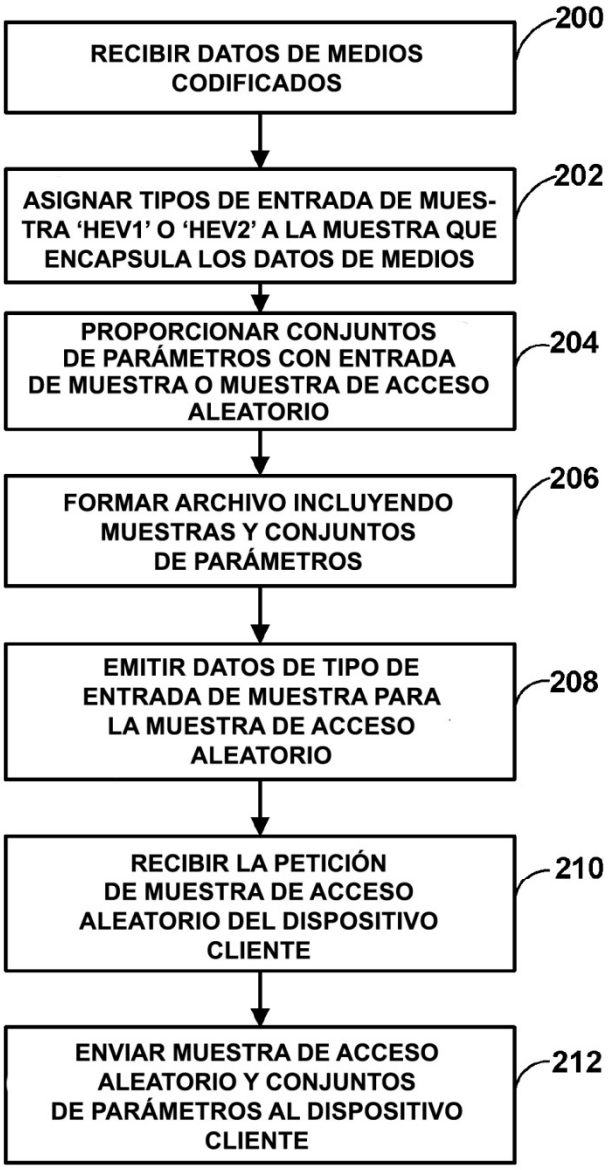
RECIBIR DATOS DE MEDIOS CODIFICADOS
200
ASIGNAR TIPOS DE ENTRADA DE MUESTRA 'HEV1' O 'HEV2' A LA MUESTRA QUE ENCAPSULA LOS DATOS DE MEDIOS
202
PROPORCIONAR CONJUNTOS DE PARÁMETROS CON ENTRADA DE MUESTRA O MUESTRA DE ACCESO ALEATORIO
204
FORMAR ARCHIVO INCLUYENDO MUESTRAS Y CONJUNTOS DE PARÁMETROS
206
EMITIR DATOS DE TIPO DE ENTRADA DE MUESTRA PARA LA MUESTRA DE ACCESO ALEATORIO
208
RECIBIR LA PETICIÓN DE MUESTRA DE ACCESO ALEATORIO DEL DISPOSITIVO CLIENTE
210
ENVIAR MUESTRA DE ACCESO ALEATORIO Y CONJUNTOS DE PARÁMETROS AL DISPOSITIVO CLIENTE
212

FIG. 5

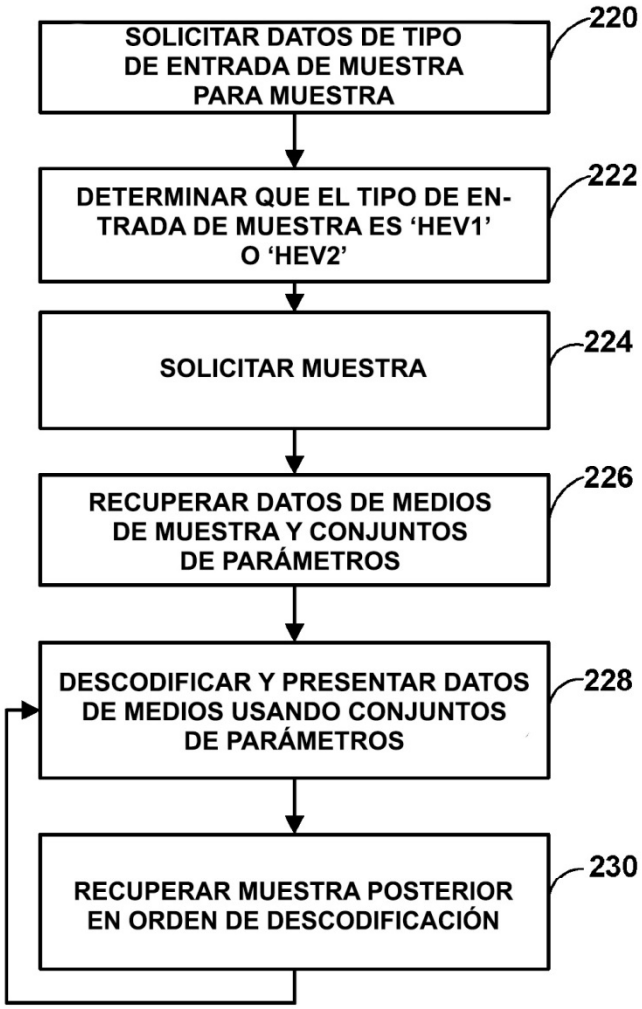
SOLICITAR DATOS DE TIPO DE ENTRADA DE MUESTRA PARA MUESTRA
220
DETERMINAR QUE EL TIPO DE ENTRADA DE MUESTRA ES 'HEV1' O 'HEV2'
222
SOLICITAR MUESTRA
224
RECUPERAR DATOS DE MEDIOS DE MUESTRA Y CONJUNTOS DE PARÁMETROS
226
DESCODIFICAR Y PRESENTAR DATOS DE MEDIOS USANDO CONJUNTOS DE PARÁMETROS
228
RECUPERAR MUESTRA POSTERIOR EN ORDEN DE DESCODIFICACIÓN
230

FIG. 6

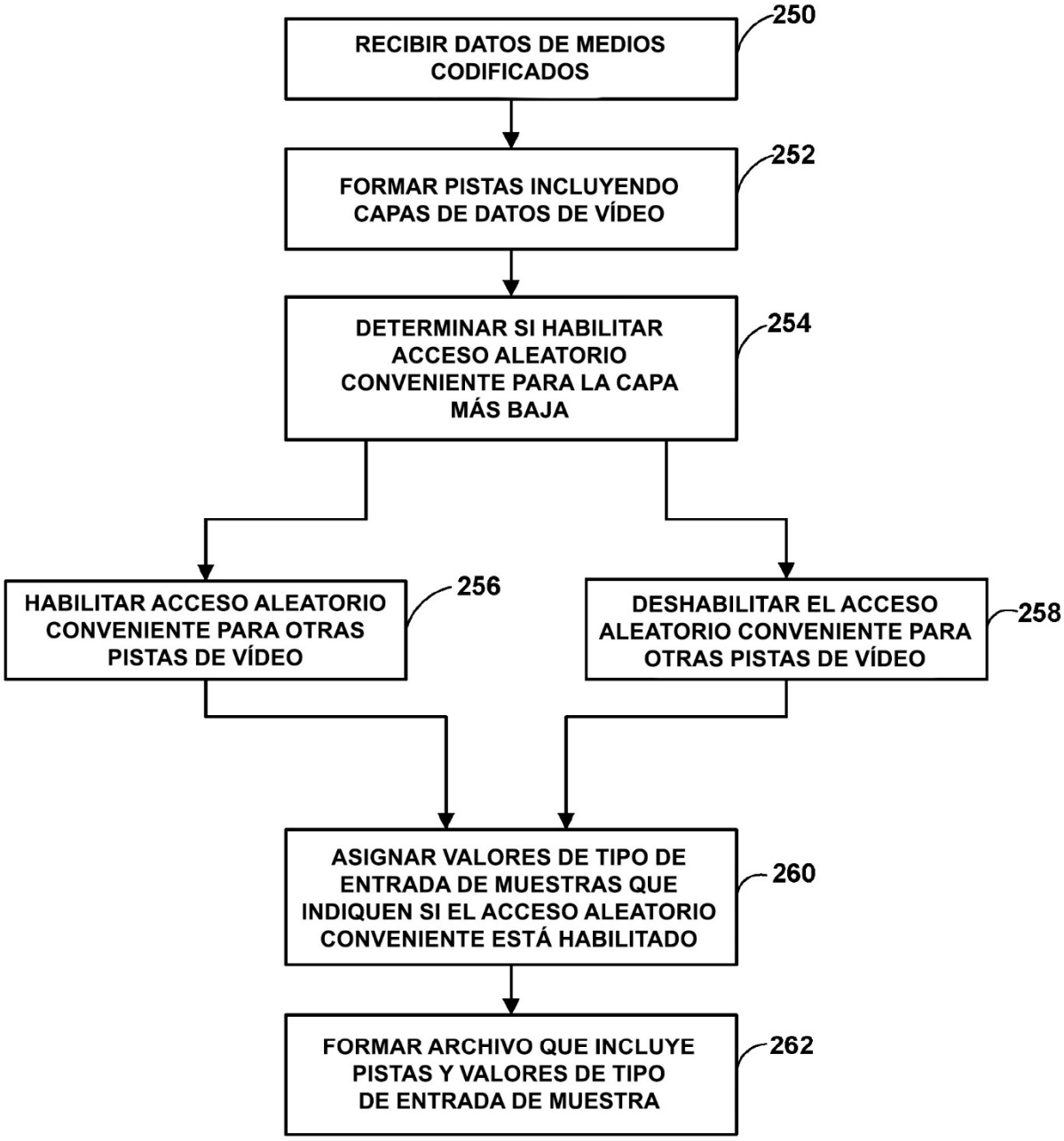
RECIBIR DATOS DE MEDIOS CODIFICADOS
250
FORMAR PISTAS INCLUYENDO CAPAS DE DATOS DE VÍDEO
252
DETERMINAR SI HABILITAR ACCESO ALEATORIO CONVENIENTE PARA LA CAPA MÁS BAJA
254
HABILITAR ACCESO ALEATORIO CONVENIENTE PARA OTRAS PISTAS DE VÍDEO
256
DESHABILITAR EL ACCESO ALEATORIO CONVENIENTE PARA OTRAS PISTAS DE VÍDEO
258
ASIGNAR VALORES DE TIPO DE ENTRADA DE MUESTRAS QUE INDIQUEN SI EL ACCESO ALEATORIO CONVENIENTE ESTÁ HABILITADO
260
FORMAR ARCHIVO QUE INCLUYE PISTAS Y VALORES DE TIPO DE ENTRADA DE MUESTRA
262

FIG. 7